Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E ALIMENTARI

Ciclo 36

Settore Concorsuale: 07/G1 - SCIENZE E TECNOLOGIE ANIMALI

Settore Scientifico Disciplinare: AGR/17 - ZOOTECNICA GENERALE E MIGLIORAMENTO GENETICO

APPLIED GENOMICS IN LIVESTOCK: GENOME-WIDE ASSOCIATION STUDIES AND POPULATION GENOMIC ANALYSES IN PIG AND RABBIT BREEDS

Presentata da: Mohamad Ballan

Coordinatore Dottorato

Massimiliano Petracci

Supervisore

Luca Fontanesi

Co-supervisore

Stefania Dall'Olio

Esame finale anno 2024

Acknowledgments

I would like to express my sincere gratitude and appreciation to my supervisor **Prof. Luca Fontanesi**, for the continuous support, for his patience, motivation, immense knowledge, and for the valuable experiences that he provided me to achieve during these three years. Without your influential support, feedback, and encouragement this accomplishment would have never been achieved.

Besides, I would like to thank in particular **Prof. Stefania Dall'Olio**, **Dr. Samuele Bovo**, **Dr. Giuseppina Schiavo**, for their encouragement, insightful comments, and the help to do this thesis and obtaining new knowledge as well as their worthwhile contribution and help in the data analysis and laboratory procedures.

I would thank **ANAS staff** for their sincere collaboration in samples and data collection during this project. I would like to thank all **my colleagues** with which I've had the opportunity to learn something from them and for all of their kind supports during my PhD path.

Abstract

The domestication and selection processes in pigs and rabbits have resulted in the constitution of multiple breeds with broad phenotypic diversity. Population genomics analysis and Genome-wide association study analysis can be utilized to gain insights into the ancestral origins, genetic diversity, and the presence of lethal mutations across these diverse breeds. In this thesis, we analysed the dataset obtained from three Italian Pig breeds to detect deleterious alleles. We screened the dataset for genetic markers showing homozygous deficiency using two approaches single marker and haplotype-based approach. Moreover, Genome-wide association study analyses were performed to detect genetic markers associated with pigs' reproductive traits. In rabbits, we investigated the application of SNP bead chip for detection signatures of selection in rabbits using different methods. This analysis was implemented for the first time in different fancy and meet rabbit breeds. Multiple approaches were utilized for the detection of the selection of signatures including F^{st} analysis, ROH analysis, *PCAdapt* analysis, and haplotype-based analysis. The analysis in pigs was able to identify five putative deleterious SNPs and nine putative deleterious haplotypes in the three analysed Italian Pig breeds. The genomic regions of the detected putative deleterious genomic markers harboring loss of function variants such as the Frameshift variant, start lost, and splice donor variant. Those variants are close to important candidate genes such as *IGF2BP1, ADGRL4, POMT2,* and *HGF.* In rabbits, multiple genomic regions were detected to be under selection of signature. These genomic regions harbor candidate genes associated with coat color phenotype (*MC1R*, *TYR*, *OCA2,* and *ASIP*), hair structure (*LIPH*), and body size (*HMGA2* and *COL2A1*). The described results in rabbits and pigs could be used to improve breeding programs by excluding the deleterious genetic markers carriers and incorporating candidate genes for coat color, body size, and meat production in rabbit breeding programs to enhance desired traits.

Table of Contents

General introduction

Pig genome mapping and sequencing

The pig genome consists of 18 autosomes, two sex chromosomes (chrX : NC_010461.5 and chrY: NC_010462.3), and Mitochondrial DNA with an estimated size of 2.5 Gb. In comparison to humans, there is a similarity between the pig genome and human genomes. The N50 size represents the point where half of the genome assembly consists of blocks that are as long as or longer than the N50 size. For the contigs of Pig genome, the N50 length is 48,231,277, and for the scaffolds, it's 88,231,837. The pig genome has been sequenced by the Swine Genome Sequencing Consortium (SGSC) using a hybrid approach, which involves combining hierarchical shotgun sequencing of BAC clones with whole genome sequencing (Archibald et al. 2010), by using the 3x coverage of the BACS and the 3x coverage of the whole genome approach. The revised assembly (Sscrofa Build 10) has been released and is accessible for use by the scientific community. Sscrofa11 (GCA_000003025.6) is the recently updated assembly of S. scrofa genome that was released by SGSC in 2017. To construct this assembly, the sequence data primarily consisted of 65x genome coverage using whole genome shotgun (WGS) Pacific Biosciences long reads (Pacific Biosciences RSII, with P6/C4 chemistry). Illumina HiSeq2500 WGS paired-end and mate-pair reads were used for final error correction. Additionally, Sanger and Oxford Nanopore sequence data from a few CHORI-242 BAC clones were utilized to fill gaps. All WGS data were derived from a single Duroc female (TJ Tabasco, or Duroc 2-14) (Warr et al. 2020). The total genome size of the recent genome assembly sus. scrofa 11 is 2.5 GB. The number of annotated genes is 30,371 genes including 22,063 protein-coding genes and 13,145 protein non-coding genes. The PorcineSNP60 BeadChip is a genotyping array designed explicitly for the porcine genome. It provides capabilities, for analysing variations in various pig breeds, such as Duroc,

Landrace, Pietrain, and Large White. This SNP BeadChip consists of 64,232 SNPs that cover the entire porcine genome, which offers a wide array of applications. These include genome-wide selection, identifying quantitative trait loci, assessing genetic merit, cross-breed mapping, conducting linkage disequilibrium studies, undertaking comparative genetic research, copy number variant detection (CNV), the detection of unfavorable alleles, and characterizing breeds to evaluate biodiversity.

Domestication and selection processes in pigs

The pig is considered one of the first domesticated animals and an essential source of animal protein for human nutrition (Boland et al. 2013). The domestication of pigs has influenced the advancement of agriculture and the adoption of farming practices around the world. The first evidence of the domestication process in pigs has been discovered at Çayönü in southeastern Anatolia, dating back to 7000 BC (Ervynck et al. 2001). The domestic pig likely originated in the Near East and then expanded its presence westward into Europe and eastward into China. However, recent research using mitochondrial DNA (mtDNA) sequencing indicates that S. scrofa species actually originated from islands in Southeast Asia, such as the Philippines and Indonesia (Giuffra et al. 2000). Domestic pigs exist in various global environments with multiple phenotypes, such as teeth and skull shape, external proportions, hair and colour patterns, biochemical and molecular differences, ecology, behavior, reproductive isolation, and natural habitats. Those phenotypes are used to differentiate the genus of sus. Scrofa to different subspecies, with over 16 distinct subspecies identified and each occupying specific geographic regions.

The pig is considered a domesticated animal with several closely related outgroup subspecies and surviving wild subspecies, making it highly suitable for studying ancestral mutations and the impact of selective processes. Ancestral mutations are significant because they are linked to allele frequency and regions affected by positive selection. it is still unclear as to what constitutes the nearest living relative (likely S. barbatus) and the age of the species *S.scorfa* relative to some of its nearest relatives.

Therefore, It is essential to compare it with various related species, dating from 1 to 6 million years ago, to understand its evolutionary history (Randi et al., 1996; Jones et al. 1998)

The artificial selection in pig domestication has led to rapid changes in physical traits such as high modification in morphological architectures and reduction in skeletal size. Moreover, domestication and selection have positively impacted the biochemical composition and other traits, bringing out significant and beneficial improvements (Groenen et al., 2016). Pig breeds in Europe and Asia exhibit wide different morphological features such as variation in body size, coat colour, and ear shape. Other reproduction and behavior phenotypes related to these pig breeds have significantly changed compared to their wild ancestors. The selection process in pigs occurs on genes within the same pathways that affect the same biological processes in other species. Therefore, Many of the morphological traits and behavioral changes that have been seen in pigs can also be observed in other domesticated species. Several of these morphological traits like coat colour, floppy ears, and curly tails, have been proposed to be directly linked to domestication (Price et al. 1999; Trut et al. 2009).

Probably, a number of these traits were preferred by certain breeders and underwent intense selection in specific pig breeds following the initial domestication, especially for coat colour. Moreover, many genes have been detected that clearly have been under strong selection such as (*KIT*,*MC1R*) for coat colour (Andersson et al. 2011; Fang et al. 2011) and *IGF2*, *RYR1* for lean growth (Van Laere et al. 2003; Fujii et al. 1991).

Italian pigs industry

The primary focus of the Italian pig industry is to produce heavy carcasses, which are used to obtain raw meat for traditional processed products, such as dry-cured hams with the Protected Designation of Origin (PDO) label. Italian heavy pigs are typically slaughtered at an average live weight of 160 kg (with a variation of ± 10%) and are at least 9 months old when processed. This period is crucial for achieving raw hams with the appropriate weight, muscle maturity, and

subcutaneous fat thickness (15–30 mm), which helps minimize seasoning losses during the processing period (Dall'Olio et al. 2020). Dry-cured ham production is carried out according to strict regulations set by the ham Consortia, including everything from breeding strategies to post-slaughter storage. The geographical origin of the animals is also carefully controlled and regulated. For animals destined for dry-cured ham production, the economic value is mainly associated with the legs that will undergo processing, and their fat content plays a crucial role in the aging process. High-quality ham does not contain additives and preservatives, and the aging is performed only by controlling the amount of added salt, humidity, and temperature of the place in which they are processed (Reg. UE n. 1151/2012, Disciplinari Prosciutto di Parma DOP, Disciplinari Prosciutto San Daniele DOP). The process that can lead to an optimal dry-cured ham begins with genetic selection. According to the National Pig Breeders Association (ANAS) programs animals are selected to improve feeding efficiency, performance and carcass traits and meat quality traits. Animals are chosen to enhance feeding efficiency, performance, carcass traits, and meat quality. Piglets are tested at 30-45 days, with regular measurements of parameters like Average Daily Gain (ADG) and feeding-to-gain ratio during their growth. After slaughtering, Back Fat Thickness (BFT) is measured, indirectly indicating the fat covering the legs. These measurements contribute to calculating Estimated Breeding Values (EBV) using the multiple traits of the mixed model known as the BLUP animal model. The Italian Large White Pigs national selection program aims to maintain traditional carcass fat coverage and enhance meat quality for typical aged products. The breeding strategy involves sib-testing, calculating Estimated Breeding Values (EBV) based on production trait measurements from triplets (2 females and 1 castrated male) of the same litter, slaughtered at around 160 kg. Integrating genomic data with simple phenotypic aspects is crucial for optimizing future breeding selection strategies. The Italian Large White, Landrace, and Duroc breeds serve as the main breeds for the primary Protected Denomination of Origin (PDO) and Protected Geographical Indication (PGI) processed pig meat products. The selection process considers various characteristics, including

the weight loss of hams during the initial salting (CALO), the presence of visible intermuscular fat in hams (VIF), backfat thickness (BFT), the weight of loins at the time of slaughter (LC), the average daily gain (ADG), the total number of live piglets born during the first farrowing (NBA), and the number of litters per sow throughout her career. Additionally, since 2022, two new traits have been included: resistance to respiratory disease (RMR) and resistance to enteric disease (RME). The main challenge lies in developing methods that can genuinely verify the authenticity of single-breed products (https://www.anas.it/).

Introduction to deleterious alleles

Lethal recessive alleles can cause death in the homozygous state affecting individuals before or after birth, leading to reduced fertility in different populations (Cole et. 2018). Normally, these recessive lethal alleles are widespread but have a minimal effect due to their very low individual mutation frequency. However, in small domestic and wild populations, inbreeding can expose these alleles when related parents with a common ancestor pass on the same recessive lethal allele (Tarsk et al. 2016; Bosse et al. 2018). The impact of recessive lethal varies based on population structure (such as effective population size) and the rate of recessive lethal mutations. In livestock, intense genomic selection has led to relatively small effective population sizes (Hall et al. 2016). With smaller effective population sizes, genetic drift can quickly increase the frequency of recessive lethal in the population. While genomic selection has significantly improved various traits like production, fertility, and disease resistance (Gonzalez et al. 2015), it may not provide substantial advantages in controlling the frequency of recessive lethal mutations compared to traditional selection methods (Dalton et al. 2015).

Each individual's genome carries harmful mutations that can affect their fitness and health. This mutational load, which depends on factors like mutation rate, demographic history, and selection, can increase the frequency of these deleterious mutations. Most harmful mutations are recessive, meaning their negative effects become apparent only in the homozygous state. Inbreeding, the inheritance of

identical genetic material from related parents, can lead to long homozygous regions in the offspring's genome, causing potential negative impacts on health and reproduction compared to outbred populations. This effect due to inbreeding is referred to as "genetic load," mainly resulting from the expression of recessive harmful mutations (Bosse et al. 2019).

In population genetics, the balance between harmful alleles emerging from mutations and those eliminated by natural selection is known as mutation-selection balance. This equilibrium maintains harmful alleles at low frequencies in the population, as purifying selection removes them while new variants arise. Recent studies have described several harmful alleles existing at moderate to high frequencies in the population (Georges et al., 2019). Such high frequencies can be the result of genetic drift and may result from another factor: The positive selection for heterozygous advantage. The heterozygous advantage depends on an interplay between the benefits of heterozygotes and the negative consequences of homozygotes, leading to a balance between purifying selection against the mutant homozygotes, and positive selection on heterozygous.

Sequencing technologies allow the estimation of the mutational load in an individual's genome from sequence data. It is important to note that the number of harmful mutations may not necessarily differ between inbred and outbred individuals for differences in fitness to occur. The crucial factor is that harmful mutations generally have a small fitness effect when in the heterozygous state, but in the homozygous state, they are expressed and may cause heritable diseases. Runs of homozygosity (ROH) regions when both haplotypes are identical by descent (IBD), leading to a higher probability of a recessive deleterious allele with a low frequency (p) to become homozygous within these regions compared to outside IBD regions (Szpiech et al. 2013).

A common approach to detect potentially deleterious alleles is screening populations for the absence of specific gene variations in a homozygous state. If populations have individuals carrying these variations in a heterozygous state, we would expect to find homozygous individuals based on the frequencies of these gene alleles and mating between carriers. This approach has been implemented

by researchers such as Derks et al. (2017), Pausch et al. (2015), and VanRaden et al. (2011). This approach relies on the biological consequences of a variant's lethality, leading to the absence of the allele in its homozygous state. Large sample size is often necessary to identify such alleles and large panels of genotyped animals. The reverse genetic screening approach is an alternative technique to detecting potentially deleterious alleles by predicting those alleles from the functionality of mutations. Next-generation sequencing has opened the possibility of precisely identifying potentially harmful mutations within individual genomes (Henn et al. 2015). The harmful impact of a variant can be predicted by examining its effects on gene function, such as alterations to proteins stop-gain, stop-loss, frameshift, etc. (Derks et al. 2019).

Charlier et al. (2016) used a reverse genetic screen to demonstrate that loss-of-function mutations and harmful missense mutations lead to embryonic lethality in cattle populations. However, identifying recessive embryonic lethal is frequently challenging due to the lack of affected individuals and their relatively low frequency. To overcome this challenge, genotyping and sequencing large cohorts of animals within a population can support the discovery of such lethal variations and directly point to the causal mutations.

Deleterious alleles detected in cattle, pigs, and other species

Vanraden et al. (2011), conducted a study with the aim of identifying recessive lethal alleles in three cattle breeds. The study examined a total of 58,453 Holsteins, 5,288 Jerseys, and 1,991 Brown cattle breeds. They performed genotyping on the samples using the BovineSNP50 BeadChip from Illumina. Five new recessive defects were discovered in Holsteins, Jerseys, and Brown Swiss breeds. These lethal alleles were identified by analyzing haplotypes with a high frequency in the population but were never found in a homozygous state. The study's results suggested that these deleterious lethal effects could impact conception, gestation, and stillbirth losses.

Sahana et al. (2013) conducted a study involving 7,937 Nordic Holstein animals to identify haplotypes showing for absence of homozygotes states. This study described 17 homozygotes deficient haplotypes, which were grouped into eight genomic regions that potentially carried recessive lethal alleles. The study assessed the impact of these identified haplotypes on two fertility traits: non-return rates and calving intervals. Among the eight genomic regions, it was confirmed that six of them indeed had an impact on fertility.

In the Belgian Blue breed, A missense variant in the *KITLG* gene has been identified leading to sterility in affected females due to the absence of Müllerian ducts. This condition is commonly referred to as the White Heifer disease, as around 90% of the affected animals exhibit a completely white phenotype (Reissmann and Ludwig, 2013). Heterozygous carrying this variant displays a 'roan phenotype', characterized by a mix of colours with some white spotting.

Kadri et al. (2014) discovered a deletion involving the *RNASEH2B* gene, which leads to embryonic lethality in homozygous mutant Nordic Red cattle. However, carriers of this deletion exhibited favorable effects on milk yield and composition. As a consequence, the carrier frequency was found to be 13%, 23%, and 32% in Danish, Swedish, and Finnish Red cattle, respectively. This high carrier frequency has been linked to a decline in fertility within these breeds in recent years (Kadri et al., 2014).

Häfliger et al. (2021) conducted a study using the reveres genetic approach to identify putative lethal alleles in the local Simmental cattle population. Within this research, 11 genomic regions were detected to exhibit homozygous deficiency. The harmful mutations detected within these 11 genomic regions affected important coding genes (*DIS3*, *CYP2B6,* and *NUBPL*) that have been suggested to be associated with embryonic lethality and unknown recessive disorders.

In Nellore cattle breed, the study of Schmidt et al. (2023) aimed to identify lethal recessive haplotypes by screening the expected population homozygosity. Moreover, the study tested the SNP markers located close to the genomic regions of lethal haplotypes for potential association with phenotypes

related to reproductive traits. This study has identified thirty potentially harmful haplotypes, and the alleles found within these haplotypes were shown to negatively impact reproductive performance, as revealed by GWAS analysis. Furthermore, the study reported multiple candidate genes that have been annotated to the genomic regions of lethal haplotypes such as *MTUS2, NFIA, A2M, A2ML1, KLRG1*.

Häggman et al. (2017) conducted a study on Finnish Yorkshire pigs to identify genomic regions with potential deleterious haplotypes. They then analyzed how these haplotypes might impact reproductive traits (Number of stillborn piglets and total number of born piglets) through GWAS analysis. The study revealed 26 putative deleterious haplotypes, with one of them significantly associated with the analyzed reproductive traits. Furthermore, three of the candidates' genes have been annotated to the genomic regions of the putative deleterious haplotype that is associated with reproductive traits *MAD2LI*, *FGF2,* and *ANXA5.*

The study of Derks et al. (2018) on the Large White pig breed identified a recessive lethal deletion within the *BBS9* gene that causes fetal lethality in mutant homozygotes. This particular lethal mutation seems to have pleiotropic effects on fertility and growth, as it leads to an increase in feed intake and growth in heterozygotes. Furthermore, the *BBS9* carrier frequency of 10.8% within this population suggests that the gene is under balancing selection. The fetal lethality observed in this pig population is a result of a 212-kb deletion in the Bardet-Biedl Syndrome 9 (*BBS9*) gene. This deletion causes a frameshift, leading to a truncated protein due to a premature stop codon. Consequently, the truncated *BBS9* protein loses its function, resulting in an increased growth rate in heterozygous pigs.

Derks et al. (2019) performed a screening of a dataset consisting of 28,085 and 11,255 animals from the Norwegian Landrace pigs and Duroc pigs purebreds, respectively. This study aimed to identify deleterious alleles within these two pig populations. The genotyping was performed using the (Illumina) Geneseek custom 50K SNP chip, which includes 50,689 SNPs (50K). The results of the study revealed five recessive lethal haplotypes, with carrier frequencies reaching up to 13.4%. Additionally, the study detected two splice-site variants affecting the *POLR1B* and *TADA2A* genes,

one frameshift variant in *URB1* gene, and one missense variant in *PNKP* gene. These variants resulted in a complete loss of function of those essential genes. Furthermore, According to this study's results, the recessive lethal alleles affect up to 2.9% of the litters within a single population and are responsible for the death of 0.52% of the total population of embryos.

In Horses, heterozygous animals for variant in the *TRPM1* gene variant display a desirable leopard complex spotting phenotype, favored by breeders. However, homozygous mutants suffer from congenital stationary night blindness. The leopard complex spotting in horses is a result of a 1378-bp insertion in TRPM1 gene's intron 1, caused by a long terminal repeat (LTR) of an endogenous retrovirus (ERV). This insertion disrupts *TRPM1* expression, leading to the prevention of translation of the last amino acids through premature poly-adenylation (Bellone et al., 2013).

Todd et al. (2020) implemented a study for screening the Thoroughbred horse population to identify harmful genetic variations. The study revealed a particular haplotype in the genomic region of the *LY49B* gene that demonstrated strong evidence of being lethal when present in a homozygous state, even though it had a high frequency in the heterozygous state. Further analysis of the whole-genome sequencing dataset identified two SNPs located in the 3′UTR of the *LY49B* gene, which could potentially lead to a loss of function. Analysing transcriptomic data from equine embryonic tissue shows that *LY49B* is expressed in the trophoblast during the placentation stage of development. these findings suggest that *LY49B* may have an essential but as yet unknown function in the implantation stage of equine development.

In rabbits, The *KIT* gene is associated with English spotted coat colour in the heterozygous state *EN/en*. The English spotting coat colour locus in rabbits, also known as the Dominant white spotting locus, is determined by an incompletely dominant allele (*En*). Rabbits with two copies of the wild-type allele (*en/en*) have a self-coloured coat, while those with one copy of the dominant allele (*En/en*) have spotted coats, and those with two copies of the dominant allele (*En/En*) appear almost

completely white but may suffer from digestive problems due to dilated cecum and ascending colon, making them subvital (Fontanesi et al., 2014).

The dwarf allele is recessive and lethal in rabbits, leading to the death of homozygous dwarf rabbits shortly after birth. Heterozygous rabbits express the dwarf phenotype, while rabbits from dwarf breeds with two copies of the wild-type allele are normal but smaller than other breeds. According to Carneiro et al. (2017), a 12.1 kb deletion in the *HMGA2* gene causes dwarfism in heterozygous individuals but results in a lethal phenotype in homozygous individuals.

Rabbit genome and domestication

The European rabbit (Oryctolagus cuniculus), also known as the "rabbit," is a unique domestic species with multiple uses such as meat and fur production. For these purposes, many commercial rabbit lines and populations have been selected. Rabbit (*Oryctolagus cuniculus*) was one of 24 mammals that has been sequenced as part of the Mammalian genome project founded by the National Institute of Health. The first release of the rabbit genome is a low-coverage 2X assembly that has been sequenced and assembled by The Broad Institute. The N50 size is the length such that 50% of the assembled genome lies in blocks of the N50 size or longer. The N50 length for super contigs is 54 kb and is 3.2 kb for contigs. The total number of bases in super contigs is 3.45 Gb and in contigs is 2.08 Gb. The OryCun2.0 is the current assembly of the rabbit genome (GCA_000003625.1) that was submitted by The Broad Institute in October 2009. The assembly consists of 84,024 contigs assembled into 3,318 scaffolds. From these sequences, 22 chromosomes have been built. The N50 length for the contigs is 64,648 while the scaffold N50 is 35,972,871. The total genome size of the recent rabbit genome assembly (OryCun2.0) is 2.7 Gb. The number of annotated genes is 20,612 coding genes and 8,319 noncoding genes.

European rabbit domestication is considered as recent process compared to the domestication of other livestock species. Oryctolagus cuniculus algirus and Oryctolagus cuniculus cuniculus are two

subspecies of European wild rabbits that have been distinguished through both morphometric analysis and molecular markers. O. c. algirus is primarily found in the southeastern region of the Iberian Peninsula, whereas O. c. cuniculus originally spread in the northeastern part of the Iberian Peninsula and the southern areas of France (reviewed in Fontanesi., 2021). The observed phylogeographic pattern in these wild European rabbit populations indicates a significant period of geographic isolation, resulting in separate evolutionary paths. This geographic separation may have been influenced by climatic barriers that emerged during the Quaternary ice ages in the Iberian Peninsula. The process of domestication initially targeted the O. c. cuniculus subspecies, which initially colonized France and later spread to other northern European regions, eventually leading to the development of multiple rabbit breeds distinguished by their external phenotypes (Zeder et al., 2006). Many of these rabbit breeds are primarily differentiated by various coat colors and patterns, which were selectively stabilized by breeders.

Several coat color genes have been described in rabbits, confirming Mendelian inheritance and homology across species. At the *extension* locus, mutations in the *MC1R* gene create three altered alleles (E^D or E^S, resulting in dominant black/steel coats; e^J, causing the Japanese brindling pattern; *e*, leading to recessive yellow/red coats) (Fontanesi et al. 2006). Mutations in the *ASIP* gene at the *agouti* locus produce black non-agouti (*a*) and tan (a^t) alleles (Fontanesi et al. 2010). Mutations in the *TYR* gene at the *albino* locus result in chinchilla, Himalayan (in the Californian breed), and full albino coat colors (Utzeri et al. 2021). In the *English spotting* locus, a marker in the *KIT* gene is associated with the spotted pattern of the Checkered Giant and Rhinelander breeds, whose classical spotted design is due to the heterozygous genotype *En/en* that preserves these animals from a megacolon defect associated with the *En* allele.

Furthermore, breeds can be differentiated by their hair structure. For instance, a mutaion in the *lipase member H (LIPH*) gene is responsible for the soft down-hair found in the Rex breed (Diribarne et al. 2011). Differences in ear shape and position also set apart certain breeds, like various lop breeds. Additionally, rabbit breeds are categorized based on their adult body weight, falling into dwarf, small,

medium, or large. For instance, Carneiro et al. (2017) described a significant deletion in the *HMGA2* gene as the cause of one type of dwarfism in rabbits.

Genome wide assosciation studies and signature of selection analysis in rabbits.

A new high-throughput genotyping tool was developed recently for rabbits. The genotyping array, known as the Orcun SNP array, utilizes Affymetrix's axiom genotyping platform and includes 199,692 SNPs. It has proven invaluable in several genome-wide association studies that have pinpointed specific genomic regions influencing the quality of meat and reproduction traits within different rabbit populations. This tool has been used in a few genome-wide association studies that detected genomic regions affecting meat quality traits and reproduction traits in meat rabbit lines. Casto-Rebollo et al . (2020) utlilized two GWAS approaches (single-marker regression and Bayesian multiple-marker regression), with the aim of detecting genes and functional mutations linked to the variability in Environmental variance (V_E) of litter size (LS) in rabbits. Their investigation yielded findings that identified four specific genomic regions, including a total of 38 genes that were strongly associated with V_E of LS. These genes were found to be related to the immune system, the development of sensory structures, and responses to stress. Sosa Madrid et . (2020) conducted a study using a GWAS analysis to identify the genomic regions linked to Intramuscular fat (IMF) in rabbits. The findings of this investigation provide strong evidence that four distinct genomic regions (OCU1, OCU8, and OCU13) have a significant impact on IMF levels in rabbits. Furthermore, the study highlighted several important candidate genes, including *EWSR1*, *MTMR2*, *APOLD1*, *PLBD1*, *PDE6H*, *GPRC5A*, and *KRAS*, which are associated with IMF content in these rabbits. Another example of GWAS studies in rabbits is the study of Bovo et al. (2021). The study conducted GWAS approaches to identify genes associated with a number of teats in 247 samples of Italian White rabbits. The GWAS identified 50 significant SNPs associated with the number of teat number traits in rabbits. The most significant SNP was close to the *NUDT2* gene, a breast carcinoma cell proliferation promoter. Another significant SNP identified is the *NR6A1* gene, which is well known to play an

important role in affecting the number of vertebrae in pigs. Furthermore, A high-throughput genotyping array can be used for cost-effective comprehensive genome analyses in rabbits. It allows for the detection of population structures and relationships among animals, families, lines, and breeds, identify signatures of selections, or infer population histories. Other applications could be for parentage analysis, traceability of rabbit products and animals, the construction of high-resolution genetic maps, and for the design of genomic selection programmes in rabbits.

References

Andersson, L., & Plastow, G. (2011). Molecular genetics of coat colour variation. In *The genetics of the pig* (pp. 38-50). Wallingford UK: CABI.

Archibald, A. L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M. A., Harlizius, B., & Swine Genome Sequencing Consortium. (2010). Pig genome sequence-analysis and publication strategy. *BMC genomics*, *11*, 1-5. https://doi.org/10.1186/1471-2164-11-438.

Bellone, R. R., Holl, H., Setaluri, V., Devi, S., Maddodi, N., Archer, S., & Brooks, S. A. (2013). Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PloS one*, *8*(10), e78280. https://doi.org/10.1371/journal.pone.0078280.

Boland, M. J., Rae, A. N., Vereijken, J. M., Meuwissen, M. P., Fischer, A. R., van Boekel, M. A.,, & Hendriks, W. H. (2013). The future supply of animal-derived protein for human consumption. *Trends in food science & technology*, *29*(1), 62-73. https://doi.org/10.1016/j.tifs.2012.07.002.

Bosse, M., Megens, H. J., Derks, M. F., de Cara, Á. M., & Groenen, M. A. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary applications*, *12*(1), 6-17. https://doi.org/10.1111/eva.12691.

Bovo, S., Schiavo, G., Utzeri, V. J., Ribani, A., Schiavitto, M., Buttazzoni, L., ... & Fontanesi, L. (2021). A genome-wide association study for the number of teats in European rabbits

(Oryctolagus cuniculus) identifies several candidate genes affecting this trait. *Animal Genetics*, *52*(2), 237-243.

Cabot, R. A., Kühholzer, B., Chan, A. W. S., Lai, L., Park, K. W., Chong, K. Y., & Prather, R. S. (2001). Transgenic pigs produced using in vitro matured oocytes infected with a retroviral vector. *Animal biotechnology*, *12*(2), 205-214. https://doi.org/10.1081/ABIO-100108347.

Carneiro, M., Hu, D., Archer, J., Feng, C., Afonso, S., Chen, C., & Andersson, L. (2017). Dwarfism and altered craniofacial development in rabbits is caused by a 12.1 kb deletion at the HMGA2 locus. *Genetics*, *205*(2), 955-965. https://doi.org/10.1534/genetics.116.196667.

Casto-Rebollo, C., Argente, M. J., García, M. L., Pena, R., & Ibáñez-Escriche, N. (2020). Identification of functional mutations associated with environmental variance of litter size in rabbits. *Genetics Selection Evolution*, *52*(1), 1-9.

Charlier, C., Li, W., Harland, C., Littlejohn, M., Coppieters, W., Creagh, F., & Georges, M. (2016). NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Research*, *26*(10), 1333-1341. https://doi.org/10.1101/gr.207076.116.

Cole, J. B., VanRaden, P. M., Null, D. J., Hutchison, J. L., Cooper, T. A., & Hubbard, S. M. (2017). Haplotype tests for recessive disorders that affect fertility and other traits. *USDA Animal Improvement Program Research Report Genomics*, 09-13.

Dall'Olio, S., Aboagye, G., Costa, L. N., Gallo, M., & Fontanesi, L. (2020). Effects of 17 performance, carcass and raw ham quality parameters on ham weight loss at first salting in heavy pigs, a meat quality indicator for the production of high quality dry-cured hams. *Meat science*, *162*, 108012. https://doi.org/10.1016/j.meatsci.2019.108012.

Derks, M. F., Gjuvsland, A. B., Bosse, M., Lopes, M. S., van Son, M., Harlizius, B., & Megens, H. J. (2019). Loss of function mutations in essential genes cause embryonic lethality in pigs. PLoS genetics, 15(3), e1008055. https://doi.org/10.1371/journal.pgen.1008055.

Derks, M. F., Lopes, M. S., Bosse, M., Madsen, O., Dibbits, B., Harlizius, B., & Megens, H. J. (2018). Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS genetics*, *14*(9), e1007661. https://doi.org/10.1371/journal.pgen.1007661.

Derks, M. F., Megens, H. J., Bosse, M., Visscher, J., Peeters, K., Bink, M. C., & Groenen, M. A. (2018). A survey of functional genomic variation in domesticated chickens. *Genetics selection evolution*, *50*(1), 1-14. https://doi.org/10.1186/s12711-018-0390-1.

Diribarne M, Mata X, Chantry-Darmon C, Vaiman A, Auvinet G, Bouet S, et al. A deletion in exon 9 of the *LIPH* gene is responsible for the rex hair coat phenotype in rabbits (*Oryctolagus cuniculus*). PLoS One. 2011;6:e19281

Ervynck, A., Dobney, K., Hongo, H., & Meadow, R. (2001). Born free? New evidence for the status of" Sus scrofa" at Neolithic Çayönü Tepesi (southeastern Anatolia, Turkey). *Paléorient*, 47-73.

Fontanesi L, Forestier L, Allain D, Scotti E, Beretti F, Deretz-Picoulet S, et al. Characterization of the rabbit agouti signaling protein (*ASIP*) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. Genomics. 2010;95:166–75.

Fontanesi L, Tazzoli M, Beretti F, Russo V. Mutations in the melanocortin 1 receptor (*MC1R*) gene are associated with coat colours in the domestic rabbit (*Oryctolagus cuniculus*). Anim Genet. 2006;37:489–93.

Fontanesi L. Rabbit genetic resources can provide several animal models to explain at the genetic level the diversity of morphological and physiological relevant traits. Appl Sci (Basel). 2021;11:373.

Fontanesi, L., Vargiolu, M., Scotti, E., Latorre, R., Faussone Pellegrini, M. S., Mazzoni, M., & De Giorgio, R. (2014). The KIT gene is associated with the English spotting coat color locus and congenital megacolon in Checkered Giant rabbits (Oryctolagus cuniculus). *PLoS One*, *9*(4), e93750. https://doi.org/10.1371/journal.pone.0093750.

Fujii, J., Otsu, K., Zorzato, F., De Leon, S., Khanna, V. K., Weiler, J. E., & MacLennan, D. H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science*, *253*(5018), 448-451. https://doi.org/10.1126/science.1862346.

Georges, M., Charlier, C., & Hayes, B. (2019). Harnessing genomic information for livestock improvement. *Nature Reviews Genetics*, *20*(3), 135-156.

Giuffra, E. J. M. H., Kijas, J. M. H., Amarger, V., Carlborg, Ö., Jeon, J. T., & Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, *154*(4), 1785-1791. https://doi.org/10.1093/genetics/154.4.1785.

González-Peña, D., Knox, R. V., MacNeil, M. D., & Rodriguez-Zas, S. L. (2015). Genetic gain and economic values of selection strategies including semen traits in three-and four-way crossbreeding systems for swine production. *Journal of animal science*, *93*(3), 879-891. https://doi.org/10.2527/jas.2014-8035.

Groenen, M. A. (2016). A decade of pig genome sequencing: a window on pig domestication and evolution. *Genetics Selection Evolution*, *48*(1), 1-9. https://doi.org/10.1186/s12711-016-0204-2.

Häfliger, I. M., Seefried, F. R., & Drögemüller, C. (2021). Reverse Genetic Screen for Deleterious Recessive Variants in the Local Simmental Cattle Population of Switzerland. *Animals*, *11*(12), 3535. https://doi.org/10.3390/ani11123535.

Häggman, J., & Uimari, P. (2017). Novel harmful recessive haplotypes for reproductive traits in pigs. *Journal of Animal Breeding and Genetics*, *134*(2), 129-135. https://doi.org/10.1111/jbg.12240.

Hall, S. J. G. (2016). Effective population sizes in cattle, sheep, horses, pigs and goats estimated from census and herdbook data. *Animal*, *10*(11), 1778-1785. https://doi.org/10.1017/S1751731116000914.

Henn, B. M., Botigue, L. R., Bustamante, C. D., Clark, A. G., & Gravel, S. (2015). Estimating the mutation load in human genomes. *Nature Reviews Genetics*, **16**(6), 333–343. https://doi.org/10.1038/nrg3931.

Jones, G. F. (1998). Genetic aspects of domestication, common breeds and their origin. *The genetics of the pig.*, 17-50.

Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., & Druet, T. (2014). A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLoS genetics*, *10*(1), e1004049. https://doi.org/10.1371/journal.pgen.1004049.

Matika, O., Robledo, D., Pong-Wong, R., Bishop, S. C., Riggio, V., Finlayson, H., & Houston, R. D. (2019). Balancing selection at a premature stop mutation in the myostatin gene underlies a recessive leg weakness syndrome in pigs. *PLoS Genetics*, *15*(1), e1007759. https://doi.org/10.1371/journal.pgen.1007759.

Pausch, H., Schwarzenbacher, H., Burgstaller, J., Flisikowski, K., Wurmser, C., Jansen, S., & Fries, R. (2015). Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics*, **16**, 312. https://doi.org/10.1186/s12864-015-1483-7.

Price OP. Behavioral development in animals undergoing domestication. Appl Anim Behav Sci. 1999;65:245–71.

Randi, E., Lucchini, V., & Diong, C. H. (1996). Evolutionary genetics of the Suiformes as reconstructed using mtDNA sequencing. *Journal of Mammalian Evolution*, *3*, 163-194. https://doi.org/10.1007/BF01454360.

Reissmann, M., & Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. In *Seminars in cell & developmental biology* (Vol. 24, No. 6-7, pp. 576-586). Academic Press. https://doi.org/10.1016/j.semcdb.2013.03.014.

Sahana, Goutam, et al. "Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle." *PLoS One* 8.12 (2013): e82909.

Schmidt, P. I., Mota, L. F. M., Fonseca, L. F. S., dos Santos Silva, D. B., Frezarim, G. B., Arikawa, L. M., & de Albuquerque, L. G. (2023). Identification of candidate lethal haplotypes and genomic association with post-natal mortality and reproductive traits in Nellore cattle. *Scientific Reports*, *13*(1), 10399. https://doi.org/10.1038/s41598-023-37586-z.

Sosa-Madrid, B. S., Hernández, P., Blasco, A., Haley, C. S., Fontanesi, L., Santacreu, M. A., ... & Ibáñez-Escriche, N. (2020). Genomic regions influencing intramuscular fat in divergently selected rabbit lines. *Animal genetics*, *51*(1), 58-69.

Szpiech, Z. A., Xu, J. S., Pemberton, T. J., Peng, W. P., Zollner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics*, **93**(1), 90–102. https://doi.org/10.1016/j.ajhg.2013.05.003.

Todd, E. T., Thomson, P. C., Hamilton, N. A., Ang, R. A., Lindgren, G., Viklund, Å., & Velie, B. D. (2020). A genome-wide scan for candidate lethal variants in Thoroughbred horses. *Scientific Reports*, *10*(1), 13153. https://doi.org/10.1371/journal.pone.0218407.

Trask, A. E., Bignal, E. M., McCracken, D. I., Monaghan, P., Piertney, S. B., & Reid, J. M. (2016). Evidence of the phenotypic expression of a lethal recessive allele under inbreeding in a wild population of conservation concern. *Journal of Animal Ecology*, *85*(4), 879-891. https://doi.org/10.1111/1365-2656.12503.

Trut, L., Oskina, I., & Kharlamova, A. (2009). Animal evolution during domestication: the domesticated fox as a model. *Bioessays*, *31*(3), 349-360. https://doi.org/10.1002/bies.200800070.

Van Laere, A. S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., & Andersson, L. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, *425*(6960), 832-836. https://doi.org/10.1038/nature02064.

VanRaden, P. M., Olson, K. M., Null, D. J., & Hutchison, J. L. (2011). Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of dairy science*, *94*(12), 6153-6161. https://doi.org/10.3168/jds.2011-4624.

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., & Archibald, A. L. (2020). An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience*, *9*(6), giaa051. https://doi.org/10.1093/gigascience/giaa051.

Zeder, M.A. Central questions in the domestication of plants and animals. *Evol. Anthropol.* **2006**, *15*, 105–117.

Aim

In this thesis, Applied genomics approaches were performed on the genomics datasets in Rabbit and Pig species. We analysed for the first time three Italian pig breeds to identify the deleterious alleles. The dataset consists of 10,183 samples genotyped using PorcineSNP60 BeadChip which includes around 60,000 SNPs. Two approaches were implemented here, the single marker approach and the haplotype-based approach. We also implemented breed comparison approach to discard putative deleterious alleles presented in more than one breed assuming the deleterious alleles are breed - specific. The dataset was phased using two different algorithms (beagle and SHAPEIT) and the overlapped haplotypes estimation has been adopted. SNPs and Haplotypes showing absence of homozygosity, expected to be homozygous in at least 4 samples, and showing deviation from hardy HWE were considered as putative deleterious. Finally, we analysed the available WGS dataset of the carriers of putative deleterious alleles for loss of function variants within the genomic region of putative lethal DNA markers. We also investigated the application of SNP beadchip for detection signature of selection in rabbits using different methods as this analysis is implemented for the first time in different fancy and meet rabbits breeds. In the first chapter, we have shown GWAS analysis for the number of teats in Italian Duroc breed as this treat is important reproductive treat and affected by the frequency of deleterious alleles in Italian Duroc breed. The deleterious alleles analysis were

descriped in chapter 2. Moreover, in chapters 3,4 and 5 we have shown the application of SNP beadchip for detection signature of selection in rabbits using different methods and pipelines.

CHAPTERS

1. Exploiting single-marker and haplotype-based genome-wide association studies to identify QTL for the number of teats in Italian Duroc pigs.
2. Single marker approach and haplotype based approach for detection of deleterious alleles from three Italian heavy pig breeds.
3. Genomic diversity and signatures of selection in meat and fancy rabbit breeds based on high-density marker data
4. Comparative analysis of genomic inbreeding parameters and runs of homozygosity islands in several fancy and meat rabbit breeds
5. Population genomic structures and signatures of selection define the genetic uniqueness of several fancy and meat rabbit breeds

Chapter 1 : **Exploiting single-marker and haplotype-based genome-wide association studies to identify QTL for the number of teats in Italian Duroc pigs**

Samuele Bovo[a^], Mohamad Ballan[a^], Giuseppina Schiavo[a], Anisa Ribani[a], Silvia Tinarelli[ab], Stefania Dall'Olio[a], Maurizio Gallo[b] and Luca Fontanesi[a*]

[a] Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale G. Fanin 46, 40127 Bologna, Italy

[b] Associazione Nazionale Allevatori Suini (ANAS), Via Nizza 53, 00198 Roma, Italy

[^]Equal contribution

*Corresponding author: Luca Fontanesi (LF)

E-mail address: luca.fontanesi@unibo.it

Published in Journal of Livestock Science. In March 2022

Doi: https://doi.org/10.1016/j.livsci.2022.104849.

Highlights

- Genome-wide association studies for the number of teats were carried out in 1162 Italian Duroc pigs
- One significant and one suggestive QTL were located on porcine chromosome 7 (*VRTN* gene region) and chromosome 10 (*FRMD4A* gene region)
- QTL for this trait identified in Italian Duroc pigs overlapped QTL regions reported in Italian Large White pigs and in a few other Duroc populations

Abstract

The number of teats is a morphological trait of high economic relevance for the pig industry. Here, aiming at dissecting the genomic architecture of this trait in the Italian Duroc pig population, we present the largest genome-wide association studies carried out in this Italian heavy pig breed thus far. A total of 1162 pigs, for which the number of teats was recorded, was genotyped with two high-throughput single nucleotide polymorphism (SNP) genotyping platforms (60K and 70K). Genome-wide association analyses were based on a single-marker approach and on a haplotype-based approach. Two quantitative trait loci (QTL) affecting the number of teats were identified. The most significant QTL, identified by the single-marker analysis and confirmed by the haplotype-based method, was located on *Sus scrofa* chromosome (SSC) 7, in the region of the *vertnin* (*VRTN*) gene. Suggestively associated markers (SNPs and haplotypes) were located on SSC10, in the region of the *FERM domain containing 4A* (*FRMD4A*) gene, the second identified QTL. These findings confirm previous results obtained in a few other Duroc populations. Overall, this study further supported the important role of variability in the *VRTN* gene region in affecting the number of teats in pig populations. Moreover, the results also indicated that this trait in the Italian Duroc breed, as in many other pig breeds, is affected by few QTL, with the contribution of many other genetic factors with small effects, following the classical theory of quantitative traits.

Keywords: genome; GWAS; heavy pig; single nucleotide polymorphism; *Sus scrofa*

1. Introduction

The number of teats is one of the most relevant morphological traits related to the reproductive performance of the sows. This trait is indirectly associated with the mothering ability, that in turn, determines the number of piglets weaned per sow and per year (Kim et al., 2005; Andersen et al., 2011). To maximize the number of weaned piglets, selection programs that have as main objective an increased litter size (a trait with low heritability) need to select, in parallel, for increased number of functional teats. Selection to improve the number of teats is, to some extent, facilitated by the medium to high heritability of the trait itself, as demonstrated by most studies in several pig breeds or lines that estimated the fraction of the genetic variance over the phenotypic variance for this morphological feature (e.g., Willham and Whatley 1962; Toro et al., 1986; McKay and Rahnefeld 1990; Borchers et al., 2002; Chalkias et al., 2013; Felleki and Lundeheim 2015; Balzani et al., 2016).

A broad variability in the number of teats (which could range from about 8 to 21 teats) has been observed both across and within pig breeds and lines (Borchers et al., 2002; Lopes et al., 2014; Verardo et al., 2016; Rohrer and Nonneman 2017; Dall'Olio et al., 2018; van Son et al., 2019). Despite being represented by discrete and countable values, the number of teats is considered a quantitative trait determined by the effect of a high number of genetic factors. Several studies have however identified a few quantitative trait loci (QTL) affecting this morphological trait in pigs. The first QTL investigations were based on reference populations (F2 and backcrosses) constructed using parental animals of different breeds or lines (some of which with divergent number of teats), including some hyper-prolific Chinese breeds (e.g., Wada et al., 2000; Hirooka et al., 2001; Rodríguez et al., 2005; Bidanel et al., 2008; Ding et al., 2009; Hernandez et al., 2014). Subsequent studies based on high density single nucleotide polymorphism (SNP) genotyping data explored the within breed variability via genome-wide association studies (GWAS). These studies confirmed the previously reported QTL regions and/or identified other novel genome regions affecting this trait

(e.g., Arakawa et al., 2015; Rohrer and Nonneman 2017; Tang et al., 2017; Lee et al., 2019; van Son et al., 2019). One of the most relevant QTL for the number of teats, confirmed by several studies, is located on *Sus scrofa* chromosome (SSC) 7 (Mikawa et al., 2007, 2011; Duijvesteijn et al., 2014; Rohrer and Nonneman 2017; Dall'Olio et al., 2018; van Son et al., 2019; Moscatelli et al., 2020). At this QTL, alleles segregating in many populations are determined by variability in the *vertnin* gene, also known as *vertebrae development associate*d gene (*VRTN*; Mikawa et al., 2011; Arakawa et al., 2015). Originally reported to have pleiotropic effects on the number of vertebrae (Mikawa *et al.*, 2011), *VRTN* encodes a novel DNA-binding transcription factor which regulates the transcription of a set of genes that harbor VRTN binding motifs and modulates somite segmentation via the Notch signaling pathway (Duan et al., 2018). Some studies in a few pig populations proposed that variability in other genes on SSC7 [*latent transforming growth factor binding protein 2* (*LTBP2*); *BRMS1 like transcriptional repressor* (*BRMS1L*); and *Fos proto-oncogene, AP-1 transcription factor subunit* (*FOS*)], close to the *VRTN* gene, are involved in affecting the number of teats and the number of vertebrae (Zhang et al., 2016; Park et al., 2018; Liu et al., 2020).

Genome-wide association studies are usually run with a model fitting one marker at a time. However, to further extract information at the genome-wide level, the use of haplotypes has been recommended to capture additional marker-phenotype associations that cannot be detected with the single-marker approach (Lorenz et al., 2010; Barendse., 2011). We recently applied haplotype-based analyses in a few GWAS in pigs and demonstrated that it is possible to improve the identification of genomic regions affecting a targeted phenotype, complementing and completing the results obtained via a single-marker approach (Bovo et al., 2019, 2020, 2021). The additional information that can be retrieved is population-dependent, as it is related to the level of linkage disequilibrium between QTL alleles and the genotyped markers, which may vary among breeds, lines and populations. An ascertainment bias derived by the SNP composition of the commercial genotyping panels used in the investigation is also another element that should be considered in this context.

We recently carried out GWAS for the number of teats in two heavy pig breeds, Italian Large White and Italian Landrace, and identified several genome regions affecting this trait, not overlapping between the two breeds (Moscatelli et al., 2020; Bovo et al., 2021). Only in Italian Large White, the *VRTN* gene region was associated with this trait whereas in Italian Landrace no significant effects of *VRTN* (or other markers of SSC7) were evidenced (Bovo et al., 2021). For the other heavy pig breed, Italian Duroc, included in the Italian selection program conducted by the Italian Pig Breeders Association (ANAS), only few information on the segregation of the *VRTN* gene alleles was reported thus far (Fontanesi et al., 2014a). This Duroc population is a close nucleus constituted in the 1990' that is under selection with the main aim to optimize the production of green legs transformed in Protected Designation of Origin (PDO) dry cured hams. Other purebred Duroc pig populations (e.g., from USA, Canada, China, Japan and The Netherlands) have been already investigated with GWAS to identify genome regions affecting the number of teats. The investigations involving these populations reported partially overlapping results in terms of QTL regions (Arakawa et al., 2015; Yang et al., 2016; Tan et al., 2017; van Son et al., 2019; Zhuang et al., 2020; Li et al., 2021).

In this study, we carried out two GWAS for the number of teats in Italian Duroc heavy pigs using single-marker and haplotype-based approaches. Obtained results were compared with the QTL information for the same trait reported in other pure-bred Duroc populations and in the other two Italian heavy pig breeds.

2. Materials and Methods

2.1. Animals

The animals included in this study were from the national selection program of heavy pig breeds that is run by ANAS. A total of 1164 Italian Duroc pigs, born in the years 1996-2018, was investigated (727 males and 437 females). Animals were included in the sib-testing evaluation

program run by ANAS, described in previous reports (Fontanesi et al., 2012, 2014b and 2015). The number of teats on these animals was routinely recorded by direct counting at the beginning of the testing period and animals having less than 12 teats were excluded from the herd book of this breed.

2.2. Genotyping and SNP quality

DNA was extracted by using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA) from blood samples routinely collected on all pigs included in the national selection program. A total of 606 animals were genotyped with the GeneSeek 70K GGP Porcine BeadChip (which includes 68516 SNPs) whereas the remaining 558 pigs were genotyped with the Illumina PorcineSNP60 BeadChip v.2 (which includes 61,565 SNPs). The two panels presented a total of 41862 shared genomic positions. Genotyping followed standard procedures based on the supplier's recommendations. DNA markers shared between the two SNP platforms were used in the analysis. BLAST+ v.2.7.1 (Camacho et al., 2009) was used to map SNPs to the Sscrofa11.1 reference genome and markers assigned to more than one position or assigned to sex chromosomes were discarded. Genotypes were evaluated with PLINK v.1.09 (Chang et al., 2015). Individual pigs with a call rate >0.90 and SNPs with a call rate > 0.90, a minor allele frequency (MAF) > 0.01 and in Hardy-Weinberg equilibrium (P > 0.0001) were retained for further analyses. The final dataset counted 1162 animals (727 males and 435 females) and 29604 DNA markers.

2.3. Haplotype estimation

The software SHAPEIT v.2 (Delaneau et al., 2011) was used for genotype phasing considering: (1) a genomic window size of 2 Mb, (2) an effective population size (N_e) estimated with SNeP v.1.1 (Barbato et al., 2015) and (3) a chromosome specific recombination rate given by Tortereau et al. (2012). The R package GHap 1.2.2 (Utsunomiya et al., 2016) was further used to call haplotypes considering a genomic window of 400 kb with a sliding block of 100 kb (Veroneze et al., 2013). Each haplotype was treated as bi-allelic DNA variant (genotypes are: NN, NH and HH; H = haplotype allele and N = NULL = all other *N* alleles present in the haploblock. PLINK was used to filter out haplotypes having a MAF < 0.02. A total of 85590 haplotypes coming from 66363

overlapping haploblocks were estimated.

2.4. Linear mixed model analyses

To assess the association between the DNA markers (SNP and haplotypes) and the number of teats we used a linear mixed effect model. An additive genetic model assuming a trend per copy of the minor allele that specify the dependency of the number of teats on genotype categories was implemented as follows:

$$\boldsymbol{y} = \boldsymbol{W\alpha} + \boldsymbol{x}\beta + \boldsymbol{g} + \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{y}$ ($n \times 1$) is a vector containing the phenotype (the number of teats) for the n^{th} animal, $\boldsymbol{W}$ ($n \times k$) is a covariate matrix with $k = 2$ (a column of 1s and a column for the sex of the animal) and $\boldsymbol{\alpha}$ is the k-dimensional vector of covariates effects, $\boldsymbol{x}$ ($n \times 1$) is the vector containing genotypes for the i^{th} DNA marker (SNP or haplotype), β is the additive fixed effect of the i^{th} DNA marker on the phenotype, $\mathbf{g} \sim N(\mathbf{0}, \sigma^2_g \mathbf{K})$ is a multivariate Gaussian polygenic effect, with covariance matrix proportional to the relatedness matrix $\mathbf{K}$ ($n \times n$) and $\mathbf{e} \sim N(\mathbf{0}, \sigma^2_e \mathbf{I})$ is a multivariate Gaussian vector of uncorrelated residuals. Polygenic effect and residuals represented the random effects. The assessment of the association between each DNA marker and the total number of teats was obtained by testing the null hypothesis $H_0: \beta = 0$. Significance was tested by using the Wald test. All the models were fitted with GEMMA v.0.98 (Zhou and Stephens 2012) after computing the relatedness matrices $\mathbf{K_1}$ and $\mathbf{K_2}$ as a centered genomic matrices, for SNPs and haplotypes, respectively. Details about the genomic matrix construction are given in the manual of GEMMA. Bonferroni correction was applied to account for multiple comparisons, considering the total number of DNA markers used in each genome scan and value of $\alpha = 0.05$. Markers presenting a $P < 5.5 \times 10^{-05}$ were considered suggestive for associations (Wellcome Trust Case Control Consortium 2007). The proportion of variance explained (PVE) by each QTL was calculated as reported in Shim et al., 2015.

SNPs and haplotypes that had the lowest *P* in chromosome regions separated by at least 5 Mb were considered as tag DNA markers. GEMMA was also used to estimate the genomic (chip) heritability (h^2_G). Genomic control inflation factor (λ_{GC}) was computed in R v.3.6.0 (R Core Team 2018). Quantile-quantile plots (QQplots) and Miami plot were generated in R by using the *qqman* package (Turner 2018).

2.5. Haploblock analysis and genome annotation

HaploView v.4.2 (Barrett et al., 2011) was used to study the structure of the haplotypes and linkage disequilibrium of the most interesting QTL regions. Each QTL was annotated considering the protein coding genes spanning the region of ± 500 kb around the evaluated DNA marker. The Sscrofa11.1 NCBI's GFF file and Bedtools v.2.17.0 (Quinlan and Hall 2010) were used for this purpose.

3. Results

Descriptive statistics

The numbers of teats in this pig population ranged from 12 to 17. About 42% of the pigs had 12 teats (Figure 1) as this represents the lower limit considered for including an animal in the Herd Book of this Italian heavy pig breed (ANAS 2021). Detailed information on the number of pigs with different numbers of teats are reported in Supplementary Table S1. On average, pigs had 12.95 ± 0.97 (mean ± standard deviation) teats. Males and females did not show any statistically significant difference in the numbers of teats ($P = 0.83$, Wilcoxon rank sum test).

Table 1 reports the number of pigs and DNA markers used in GWAS. Assessment of results via QQplots (Supplementary Figure S1) and the genomic control inflation factors (Table 1) indicated a good control of the population stratification.

Genomic heritability estimated for the number of teats in Italian Duroc pigs was moderate, with the value that increased from the estimates based on single-marker to the value obtained with the haplotype information (Table 1). This suggested that the haplotype analysis might

have captured an additional fraction of heritability, as already reported for the same trait in other pig populations (Bovo et al. 2020, 2021).

Single-marker and haplotype-based association studies

The top associated DNA markers (SNPs and haplotypes) for each QTL region are reported in Table 2, whereas Supplementary Table S2 lists all the significant and suggestively associated markers. Miami plot including SNP and haplotype data is showed in Figure 2.

Three SNPs and one haplotype on SSC7 were significantly associated with the number of teats. The most significant SNP was MARC0038565 (rs80894106; $P = 8.58\times10^{-10}$), located on SSC7 at position 97652632 bp, in the region of the *VRTN* gene (located on SSC7:97.61-97.62). This marker had a MAF = 0.44 (allele A) and a negative regression coefficient (β = -0.29), indicating a decrease in number of teats while increasing the number of copies of the minor allele. Hence, the Italian Duroc population had a higher frequency of the favorable allele (allele G) at this SNP position. The other two significant markers were closely located to the mentioned SNP and had a medium-high linkage disequilibrium (r^2 = 0.64, Figure 3) with the most significant marker. In general, this region did not show a very high level of linkage disequilibrium spanning the genotyped markers (Figure 3).

The peak of association on SSC7 was also highlighted by the haplotype analysis (SSC7:97.6-97.8 Mb; haplotype GAG; $P = 2.51\times10^{-8}$), confirming the involvement of the *VRTN* gene region in affecting the number of teats in the Italian Duroc pig breed. The DNA marker MARC0038565 was included in this haplotype [first position in the haplotype sequence AAG; Figure 3]. Hence, the haplotype GAG [higher frequent in the population with f(GAG) = 0.52; favorable allele] and marker MARC0038565 presented similar MAF and β values.

A suggestively associated chromosome region identified both using the single-marker and the haplotype-based analyses was detected in the region of the *FERM domain containing 4A*

(*FRMD4A*) gene located on SSC10:47.3-48.1 Mb. The results for the most significant SNP in this region, located at position 47947214 bp (ASGA0090802; rs81309209; $P = 4.80\times10^{-06}$), were confirmed by the haplotype analysis ($P = 5.08\times10^{-06}$). Both regions (SSC7 and SSC10) overlapped with the two QTL regions already reported in a previous study to affect the number of teats in Duroc pigs (van Son et al. 2019).

Discussion

Duroc pig populations are well known to have, in general, lower reproduction performances than other cosmopolitan pig breeds, as they are usually selected to maximize production performances and meat quality traits for the use of Duroc boars in terminal crosses (e.g., Cameron 1990; Gaugler et al. 1984; Skorupski et al. 1996; Hoque et al. 2007: Alam et al. 2021). However, to maintain a sustainable breeding in the selection nuclei of this breed, it is also needed to improve reproduction efficiency starting from traits with higher heritability.

A lower number of teats has been reported in Duroc populations than in the other two main cosmopolitan breeds, Large White and Landrace, which usually exceed the threshold of 14 teats. As such, 14 is the lower limit that is applied by ANAS to register animals to the Italian Large White and Italian Landrace herd books. In a recent survey that we carried out, the mean number of teats in these two breeds was 14.88 ± 0.92 (standard deviation) and 14.77 ± 0.87, respectively (Bovo et al. 2021), that are larger than what we observed in Italian Duroc pigs (12.95 ± 0.97). This number is however larger than what was reported in other Duroc populations. For example, Zhuang et al. (2020) reported the mean number of 10.90 ± 1.16 and 10.92 ± 1.14 in Duroc populations derived from US and Canadian nuclei and Tan et al. (2017) reported an average value of 10.72 ± 1.72 in a Chinese Duroc population. Similar mean values to what we reported in the Italian Duroc breed were reported in a Dutch Duroc line (12.93 ± 1.05; van Son et al. 2019) and in another Chinese Duroc herd (13.17 ± 1.12; Li et al. 2021). These results suggest that some heterogeneity exists in Duroc populations, probably due to different selection pressures that have been applied on

this trait in nuclei of this cosmopolitan breed around the world.

Genomic heritability that we estimated in the Italian Duroc breed was close to what we already reported in Italian Landrace pigs using both single-marker and haplotype-based approaches (h^2_G ± standard error: 0.25 ± 0.02 and 0.30 ± 0.03, respectively) and a little lower than what we estimated in Italian Large White pigs (0.30 ± 0.03 and 0.43 ± 0.04, respectively; Bovo et al 2021). Genomic heritability estimates for the number of teats were also reported in other Duroc populations. In an US nucleus, the single-marker heritability was 0.19 ± 0.02, in a Canadian population was 0.34 ± 0.03 (Zhuang et al. 2020) and in a Chinese nucleus was 0.29 ± 0.05 (Li et al. 2021). These other studies, however, did not report any estimation based on haplotypes. Genomic heritability estimated using haplotype information indicated that this approach can potentially capture additional fractions of the so called "missed heritability" that would not be disclosed using single-marker approaches (Bovo et al. 2021). Therefore, the use of haplotypes might be useful to further exploit genomic information for this purpose (Lorenz et al. 2010; Barendse 2011).

Only two QTL for the number of teats, one highlighted by significant markers and one identified by suggestively significant markers, were identified in the Italian Duroc population. Results obtained by the single-marker analysis were confirmed by the haplotype-based analysis. To continue the comparison with the results obtained in other Duroc populations, it is worth to mention that a limited number of segregating QTL for this trait has been also reported in other Duroc populations (Tan et al. 2017; van Son et al. 2019; Zhuang et al. 2020; Li et al. 2021). For example, by investigating a Dutch Duroc population, van Son et al. (2019) reported only the same two QTL regions that we identified in the Italian Duroc pigs. The larger number of pigs that was investigated in the Dutch study made it possible to reach significant results even for markers in the region of the *FRMD4A* gene that, however, remained less significantly associated than what was reported for the SNPs on SSC7, located in the region of *VRTN*. The *VRTN* region also included the most significant SNPs identified in GWAS carried out in Canadian Duroc pigs and in two Chinese Duroc

populations (Tan et al. 2017; Zhuang et al. 2020; Li et al. 2021). No significant QTL for the number of teats were however identified in an American derived nucleus of Duroc pigs (Zhuang et al. 2020). Comparison with the results obtained in a Japanese Duroc population is not possible due to the different statistical approaches that were used to detect QTL regions for the number of teats (BayesC methods; Arakawa et al. 2015). The *VRTN* region, however, emerged as one of the most significant region also in this population (Arakawa et al. 2015). A QTL pattern similar to what was identified in Italian Duroc pigs was already reported in Italian Large White pigs where, in addition to the *VRTN* and *FRMD4A* gene regions, only another significant haplotype region was identified on SSC12 (Bovo et al. 2021). The Italian Landrace population was however completely different, in terms of QTL regions for the number of teats, from both the Italian Duroc and Italian Large White breeds (Bovo et al. 2021).

These comparative analyses, in general, showed that one major QTL for the number of teats segregates in the most important cosmopolitan pig breeds and derived lines and populations (like the Italian Duroc breed). This QTL is located on SSC7 for which the most plausible causative mutation has been reported in the *VRTN* gene (Mikawa et al. 2011). In Italian Duroc pigs, the frequency of the insertion allele at the *VRTN* gene suggested to be the putative favorable allele (allele Q; Fontanesi et al. 2014a), which might increase the number of vertebrae and the number of teats (Mikawa et al. 2011), is almost equal to the frequency of the positive allele of the most significant SNP we located on SSC7 (allele G of MARC0038565). This information could indirectly suggest a very high linkage disequilibrium between the two markers. Additional studies are needed to complete the linkage disequilibrium pattern of this SSC7 genome region in Italian Duroc, including genotyping data of the *VRTN* gene which might confirm the significant results obtained for the MARC0038565 SNP. Additional studies are also needed to characterize the variability in the *FRMD4A* gene and identify the causative mutation(s) affecting this trait. This gene encodes a FERM domain-containing protein that regulates epithelial polarity by connecting ADP

ribosylation factor 6 (ARF6), which is considered an important factor involved in actin cytoskeleton dynamics and membrane trafficking (Ikenouchi and Umeda 2010). Therefore, the role of this gene in affecting the number of teats is not obvious. Thus, it will be also important to clarify the molecular mechanisms that could lead to this effect and eventually evaluate the effect of other close genes in determining the QTL located on SSC10.

Conclusions

In terms of genotyped animals, this is the largest GWAS carried out in the Italian Duroc pig breed thus far. The estimated genomic heritability that we obtained confirmed that genetic factors affect this trait in a similar way as it was evidenced in other Duroc populations. The limited number of major QTL influencing this trait and segregating in the Italian Duroc breed indicates that a marker assisted selection based on *VRTN* alleles and eventually on *FRMD4A* gene markers, might be effective but it could not completely exploit the genetic variance related to this trait for which many other genes are involved, whose small effect could only be captured including this trait in a genomic selection programs that can take into account genome-wide variability.

Acknowledgements

The authors acknowledge the collaboration of personnel of ANAS for the collaboration in this study.

Ethical approval

All animals used in this study were kept according to Italian and European legislation for pig production and all procedures described were in compliance with national and European Union regulations for animal care and slaughtering. All animals were part of the routine Italian pig breeding program and were slaughtered in a commercial authorized abattoir following standard procedures. All animals were not raised or sampled for the purpose of this study. As no treatment was given to any animals, no ethical approval was needed according to the rules of the animal research ethics committee of the University of Bologna based on the Italian legislation, as reported in the “DECRETO LEGISLATIVO 4 marzo 2014, n. 26”.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability statement.

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Funding details

This study received funds from the MIUR PRIN2017 PigPhenomics project and by the University of Bologna RFO 2019-2021 programmes. This study was also supported by the PSRN (Progetto di Sviluppo Rurale Nazionale) SUIS and SUIS2.0 projects (co-funded by the European Agricultural Fund for Rural Development of the European Union and by the Italian Ministry of Agriculture, Food and Forestry - MiPAAF).

ORCID

Samuele Bovo: 0000-0002-5712-8211

Mohamad Ballan: 0000-0003-4585-7502

Giuseppina Schiavo: 0000-0002-3497-1337

Anisa Ribani: 0000-0001-6778-1938

Stefania Dall'Olio: 0000-0003-1384-3771

Luca Fontanesi: 0000-0001-7050-3760

References

Alam, M., Chang, H-K., Lee, S-S., Choi, T-J. 2021. Genetic analysis of major production and reproduction traits of Korean Duroc, Landrace and Yorkshire Pigs. Animals 11(5):1321.

ANAS, 2019. Libro Genealogico. Retrieved on the 2nd September 2019, from http://www.anas.it/.

Andersen, IL., Nævdal, E., Bøe, KE. 2011. Maternal investment, sibling competition, and offspring survival with increasing litter size and parity in pigs (*Sus scrofa*). Behav Ecol Sociobiol. 65(6):1159–1167.

Arakawa, A., Okumura, N., Taniguchi, M., Hayashi, T., Hirose, K., Fukawa, K., Ito, T., Matsumoto, T., Uenishi, H., Mikawa, S. 2015. Genome-wide association QTL mapping for teat number in a purebred population of Duroc pigs. Anim Genet. 46(5):571–575.

Balzani, A., Cordell, HJ., Sutcliffe, E., Edwards, SA. 2016. Heritability of udder morphology and colostrum quality traits in swine. J Anim Sci. 94(9):3636–3644.

Barbato, M., Orozco-terWengel, P., Tapio, M., Bruford, MW. 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 6:109.

Barendse, W. 2011. Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. PLOS ONE. 6(12):e29601.

Barrett, JC., Fry, B., Maller, J., Daly, MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21(2):263–265.

Bidanel, JP., Rosendo, A., Iannuccelli, N., Riquet, J., Gilbert, H., Caritez, JC., Billon, Y., Amigues, Y., Prunier, A., Milan, D. 2008. Detection of quantitative trait loci for teat number and female reproductive traits in Meishan × Large White F2 pigs. Animal. 2(6):813–820.

Borchers, N., Reinsch, N., Kalm, E. 2002. Teat number, hairiness and set of ears in a Piétrain cross: variation and effects on performance traits. Arch Anim Breed. 45(5):465–480.

Bovo, S., Ballan, M., Schiavo, G., Gallo, M., Dall'Olio, S., Fontanesi, L. 2020. Haplotype-based genome-wide association studies reveal new loci for haematological and clinical-biochemical parameters in Large White pigs. Anim Genet. 51(4):601–606.

Bovo, S., Ballan, M., Schiavo, G., Ribani, A., Tinarelli, S., Utzeri, VJ., Dall'Olio, S., Gallo, M., Fontanesi, L. 2021. Single-marker and haplotype-based genome-wide association studies for the number of teats in two heavy pig breeds. Anim Genet. 52(4):440-450.

Bovo, S., Mazzoni, G., Bertolini, F., Schiavo, G., Galimberti, G., Gallo, M., Dall'Olio, S., Fontanesi, L. 2019. Genome-wide association studies for 30 haematological and blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. Sci Rep. 9(1):7003.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421.

Cameron, ND., 1990. Comparison of Duroc and British Landrace pigs and the estimation of genetic and phenotypic parameters for growth and carcass traits. Animal Science. 50(1):141–153.

Chalkias, H., Rydhmer, L., Lundeheim, N. 2013. Genetic analysis of functional and non-functional teats in a population of Yorkshire pigs. Livest. Sci. 152(2):127–134.

Chang, CC., Chow, CC., Tellier, LC., Vattikuti, S., Purcell, SM., Lee, JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4:7.

Dall'Olio, S., Ribani, A., Moscatelli, G., Zambonelli, P., Gallo, M., Nanni, Costa L., Fontanesi, L. 2018. Teat number parameters in Italian Large White pigs: phenotypic analysis and association with *vertnin* (*VRTN*) gene allele variants. Livest Sci 210:68–72.

Delaneau, O., Marchini, J., Zagury, J-F. 2011. A linear complexity phasing method for thousands of genomes. Nat Methods. 9(2):179–181.

Ding, N., Guo, Y., Knorr, C., Ma, J., Mao, H., Lan, L., Xiao, S., Ai, H., Haley, CS., Brenig, B., Huang, L. 2009. Genome-wide QTL mapping for three traits related to teat number in a White Duroc x Erhualian pig resource population. BMC Genet. 10:6.

Duan, Y., Zhang, H., Zhang, Z., Gao, J., Yang, J., Wu, Z., Fan, Y., Xing, Y., Li, L., Xiao, S., et al. 2018. VRTN is required for the development of thoracic vertebrae in mammals. Int J Biol Sci. 14(6):667–681.

Duijvesteijn, N., Veltmaat, JM., Knol, EF., Harlizius, B. 2014. High-resolution association mapping of number of teats in pigs reveals regions controlling vertebral development. BMC Genomics 15:542.

Felleki, M., Lundeheim, N. 2015. Genetic heteroscedasticity of teat count in pigs. J Anim Breed Genet. 132(5):392–398.

Fontanesi, L., Schiavo, G., Galimberti, G., Calò, DG., Scotti, E., Martelli, PL., Buttazzoni, L., Casadio, R., Russo, V. 2012. A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. BMC Genomics. 13:583.

Fontanesi, L., Scotti, E., Buttazzoni, L., Dall'Olio, S., Russo, V. 2014a. Investigation of a ShortI Nuclear Element polymorphic site in the porcine vertnin Gene: allele frequencies and association study with meat quality, carcass and production traits in Italian Large White pigs. Ital J Anim Sci 13(1):3090.

Fontanesi, L., Schiavo, G., Galimberti, G., Calò, DG., Russo, V. 2014b. A genomewide association study for average daily gain in Italian Large White pigs. J Anim Sci. 92(4):1385–1394.

Fontanesi, L., Schiavo, G., Scotti, E., Galimberti, G., Calò, DG., Samorè, AB., Gallo, M., Russo, V., Buttazzoni, L. 2015. A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. J Anim Breed Genet. 132(3):239–246.

Gaugler, HR., Buchanan, DS., Hintz, RL., Johnson, RK. 1984. Sow productivity comparisons for four breeds of swine: purebred and crossbred litters. J Anim Sci. 59(4):941–947.

Hernandez, SC., Finlayson, HA., Ashworth, CJ., Haley, CS., Archibald, AL. 2014. A genome-wide linkage analysis for reproductive traits in F2 Large White × Meishan cross gilts. Anim Genet. 45(2):191–197.

Hirooka, H., de Koning, DJ., Harlizius, B., van Arendonk, JA., Rattink, AP., Groenen, MA., Brascamp, EW., Bovenhuis, H. 2001. A whole-genome scan for quantitative trait loci affecting teat number in pigs. J Anim Sci. 79(9):2320–2326.

Hoque, MA., Suzuki, K., Kadowaki, H., Shibata, T., Oikawa, T. 2007. Genetic parameters for feed efficiency traits and their relationships with growth and carcass traits in Duroc pigs. J Anim Breed Genet. 124(3):108–116.

Ikenouchi, J., Umeda, M. 2010. FRMD4A regulates epithelial polarity by connecting Arf6 activation with the PAR complex. Proc Natl Acad Sci USA. 107(2):748–753.

Kim, JS., Jin, DI., Lee, JH., Son, DS., Lee, SH., Yi, YJ., Park, CS. 2005. Effects of teat number on litter size in gilts. Anim Reprod Sci. 90(1–2):111–116.

Lee, J., Lee, S., Park, J-E., Moon, S-H., Choi, S-W., Go, G-W., Lim, D., Kim, J-M. 2019. Genome-wide association study and genomic predictions for exterior traits in Yorkshire pigs. J Anim Sci. 97(7):2793–2802.

Li, Y., Pu, L., Shi, L., Gao, H., Zhang, P., Wang, L., Zhao, F. 2021. Revealing New Candidate Genes for Teat Number Relevant Traits in Duroc Pigs Using Genome-Wide Association Studies. Animals. 11(3):806.

Liu, Q., Yue, J., Niu, N., Liu, X., Yan, H., Zhao, F., Hou, X., Gao, H., Shi, L., Wang, Lixian., et al. 2020. Genome-wide association analysis identified *BMPR1A* as a novel candidate gene affecting the number of thoracic vertebrae in a Large White × Minzhu intercross pig population. Animals. 10(11):E2186.

Lopes, MS., Bastiaansen, JWM., Harlizius, B., Knol, EF., Bovenhuis, H. 2014. A Genome-wide association study reveals dominance effects on number of teats in pigs. PLOS ONE. 9(8):e105867.

Lorenz, AJ., Hamblin, MT., Jannink, J-L. 2010. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLOS ONE. 5(11):e14079.

McKay, RM., Rahnefeld, GW. 1990. Heritability of teat number in swine. Can J Anim Sci. 70(2):425–430.

Mikawa, S., Morozumi, T., Shimanuki, S-I., Hayashi, T., Uenishi, H., Domukai, M., Okumura, N., Awata, T. 2007. Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (*NR6A1*). Genome Res. 17(5):586–593.

Mikawa, S., Sato, S., Nii, M., Morozumi, T., Yoshioka, G., Imaeda, N., Yamaguchi, T., Hayashi, T., Awata, T. 2011. Identification of a second gene associated with variation in vertebral number in domestic pigs. BMC Genet. 12(1):5.

Moscatelli, G., Dall'Olio, S., Bovo, S., Schiavo, G., Kazemi, H., Ribani, A., Zambonelli,P., Tinarelli, S., Gallo, M., Bertolini, F., Fontanesi, L. 2020. Genome-wide association studies for the number of teats and teat asymmetry patterns in Large White pigs. Anim Genetics. 51(4):595–600.

Park, H-B., Lee, J-B., Han, S-H., Kim, S-G., Shin, H-S., Shin, M-C., Kang, Y-J., Cho, I-C. 2018. Joint linkage and linkage disequilibrium mapping reveals association of BRMS1L with total teat number in a large intercross between Landrace and Korean native pigs. Livest Science. 218:92–96.

Quinlan, AR., Hall, IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodríguez, C., Tomás, A., Alves, E., Ramirez, O., Arqué, M., Muñoz, G., Barragán, C., Varona, L., Silió, L., Amills, M., Noguera, JL. 2005. QTL mapping for teat number in an Iberian-by-Meishan pig intercross. Anim Genetics. 36(6):490–496.

Rohrer, GA., Nonneman, DJ. 2017. Genetic analysis of teat number in pigs reveals some developmental pathways independent of vertebra number and several loci which only affect a specific side. Genet Sel Evol. 49(1):4.

Shim, H., Chasman, DI., Smith, JD., Mora, S., Ridker, PM., Nickerson, DA., Krauss, RM., Stephens, M. 2015. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLOS ONE. 10(4):e0120758.

Skorupski, MT., Garrick, DJ., Blair, HT. 1996. Estimates of genetic parameters for production and reproduction traits in three breeds of pigs. New Zealand J Agric Res. 39(3):387–395.

Tan, C., Wu, Z., Ren, J., Huang, Z., Liu, D., He, X., Prakapenka, D., Zhang, R., Li, N., Da, Y., Hu, X. 2017. Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing. Genet Sel Evol. 49(1):35.

Tang, J., Zhang, Z., Yang, B., Guo, Y., Ai, H., Long, Y., Su, Y., Cui, L., Zhou, L., Wang, X., et al. 2017. Identification of loci affecting teat number by genome-wide association studies on three pig populations. Asian-Australas J Anim Sci. 30(1):1–7.

Toro, M., Dobao, MT., Rodrigáñez, J., Silio, L. 1986. Heritability of a canalized trait: teat number in Iberian pigs. Genet Sel Evol. 18(2):173–184.

Tortereau, F., Servin, B., Frantz, L., Megens, H-J., Milan, D., Rohrer, G., Wiedmann, R., Beever, J., Archibald, AL., Schook, LB., Groenen, MA. 2012. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics. 13:586.

Turner, S. 2018. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Journal of Open Source Software. 3(25):731.

Utsunomiya, YT., Milanesi, M., Utsunomiya, ATH., Ajmone-Marsan, P., Garcia, JF. 2016. GHap: an R package for genome-wide haplotyping. Bioinformatics. 32(18):2861–2862.

van Son, M., Lopes, MS., Martell, HJ., Derks, MFL., Gangsei, LE., Kongsro, J., Wass, MN., Grindflek, EH., Harlizius, B. 2019. A QTL for number of teats shows breed specific effects on number of vertebrae in pigs: bridging the gap between molecular and quantitative genetics. Front Genet. 10:272.

Verardo, LL., Silva, FF., Lopes, MS., Madsen, O., Bastiaansen, JWM., Knol, EF., Kelly, M., Varona, L., Lopes, PS., Guimarães, SEF. 2016. Revealing new candidate genes for reproductive traits in pigs: combining Bayesian GWAS and functional pathways. Genet Sel Evol. 48:9.

Veroneze, R., Lopes, PS., Guimarães, SEF., Silva, FF., Lopes, MS., Harlizius., B., Knol, EF. 2013. Linkage disequilibrium and haplotype block structure in six commercial pig lines. J Anim Sci. 91(8):3493–3501.

Wada, Y., Akita, T., Awata, T., Furukawa, T., Sugai, N., Inage, Y., Ishii, K., Ito, Y., Kobayashi, E., Kusumoto, H., et al. 2000. Quantitative trait loci (QTL) analysis in a Meishan x Göttingen cross population. Anim Genet. 31(6):376–384.

Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 447(7145):661–678.

Willham, RL., Whatley, JA. 1962. Genetic variation in nipple number in swine. Zeitschrift für Tierzüchtung und Züchtungsbiologie. 78(1–4):350–363.

Yang, J., Huang, L., Yang, M., Fan, Y., Li, L., Fang, S., Deng, W., Cui, L., Zhang, Z., Ai, H., et al. 2016. Possible introgression of the VRTN mutation increasing vertebral number, carcass length and teat number from Chinese pigs into European pigs. Sci Rep. 6:19240.

Zhang, L-C., Yue, J-W., Pu, L., Wang, L-G., Liu, X., Liang, J., Yan, H., Zhao, K-B., Li, N., Shi, H-B., et al. 2016. Genome-wide study refines the quantitative trait locus for number of ribs in a Large White × Minzhu intercross pig population and reveals a new candidate gene. Mol Genet Genomics. 291(5):1885–1890.

Zhou, X., Stephens, M. 2012. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 44(7):821–824.

Zhuang, Z., Ding, R., Peng, L., Wu, J., Ye, Y., Zhou, S., Wang, X., Quan, J., Zheng, E., Cai, G., et al. 2020. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. BMC Genomics. 21(1):344.

Tables and Figures

Table 1. Datasets used in the single-marker and haplotype-based genome-wide association studies carried out in Italian Duroc pigs.

Parameter	**Single-marker**	**Haplotype-based**
Animals used in GWAS (*n*)	1162	1162
Genotyped DNA markers used in GWAS (*n*)	29604	85590
Genomic heritability h^2_G (standard error)	0.25 (0.05)	0.30 (0.06)
Inflation factor (λ_{GC})	1.015	1.010

Table 2. Genomic regions associated and significantly associated with the number of teats in Italian Duroc pigs. Results are stratified by marker position.

Genome scan[a]	**SSC**[b]	**Pos**[c]	**Marker**[d]	**Min/Maj**[e]	**MAF**[f]	***B***[g]	***s.e.***[h]	***P***[i]	**PVE(%)**[l]	**Candidate gene**
Haplotype-based	7	97600001	CHR7_B920_GAG	N/H	0.480	-0.256	0.046	2.51×10^{-8}	2.60	*VRTN*
Single-marker	7	97652632	MARC0038565	A/G	0.442	-0.286	0.046	8.58×10^{-10}	3.22	*VRTN*
	7	98066911	H3GA0022664	A/G	0.442	-0.233	0.047	8.92×10^{-7}	2.07	*VRTN*
	7	98089286	ASGA0035527	G/A	0.442	-0.233	0.047	8.96×10^{-7}	2.07	*VRTN*
Haplotype-based	10	47000001	CHR10_B437_AGA	N/H	0.460	0.211	0.046	5.08×10^{-6}	1.78	*FRMD4A*
Single-marker	10	47947214	ASGA0090802	A/G	0.450	-0.209	0.045	4.80×10^{-6}	1.82	*FRMD4A*

[a] Genome scans performed in the Italian Duroc population. [b] *Sus scrofa* chromosome. [c] Position, in base pairs, on the *Sus scrofa* reference genome (version Sscrofa11.1). [d] DNA marker identifier reported in the chip panels. For haplotypes, the haploblock identifier (chromosome specific) and the allele is reported. [e] Minor/Major alleles. Haplotypes have been treated as bi-allelic variants (H = haplotype allele and N = other *N* alleles). [f] Minor allele frequency. [g] Regression coefficient. A positive value indicates that the no. of teats increases with the increasing of the number of copies of the minor allele. A negative value indicates that the no. of teats decreases with the increasing of the number of copies of the minor allele. [h] Standard error of the regression coefficient. [i] *P* at the Wald test of GEMMA. [l] Proportion of variance (%) explained by the DNA marker.

Figure 1. Distribution plot of the number of teats in the Italian Duroc population. The number of animals is reported at the top of each bar.

Figure 2. Miami plot of the number of teats in the Italian Duroc population. Results of the single-marker GWAS are on the top part of the plot whereas results of the haplotype-based analysis are on the bottom. Dots represent the single nucleotide polymorphism (SNP) or the haplotype markers. The red lines identify the significance thresholds (Bonferroni correction; $\alpha = 0.05$). Statistically associated SNPs are highlighted in green.

Figure 3. Regional association plot (single-marker analysis) for the SSC7:97.0-98.4 Mb genomic region. LD was measured between SNP pairs as r^2 and it is reported/showed in each box coloured in relation to its magnitude. The associated SNPs are marked with a green star symbol whereas DNA markers within the top associated haplotype (CHR7_B920_97400001_97800001_GAG) are marked with a red triangle.

Chapter 2: **Single marker and haplotype-based approaches for detection of deleterious alleles from three Italian heavy pig breeds.**

Mohamad Ballan[a], Samuele Bovo[a], Giuseppina Schiavo[a], Francesca Bertolini[a], Maurizio Gallo[b] , Manolo Cappelloni[b] and Luca Fontanesi[a*]

[a] Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale G. Fanin 46, 40127 Bologna, Italy

[b] Associazione Nazionale Allevatori Suini (ANAS), Via Nizza 53, 00198 Roma, Italy

*Corresponding author: Luca Fontanesi (LF)

E-mail address: luca.fontanesi@unibo.it

Introduction

Deleterious recessive alleles can have severe and often fatal consequences when present in the homozygous state. In livestock populations, these effects include prenatal, postnatal deaths, and reduced fertility among carrier animals (Cole et al., 2018). These harmful alleles typically have a minimal impact when they are in low frequency in the population. However, in small domestic and wild populations, inbreeding can lead to increasing their frequencies when closely related individuals with a common ancestor pass on the same recessive lethal allele (Tarsk et al., 2016; Bosse et al., 2018). Recent studies have shown the presence of lethal alleles that exist at moderate to high frequencies in the population (Georges et al., 2019). The high frequencies of these alleles can be the result of genetic drift and positive selection for heterozygous advantage. The severity of recessive lethal alleles' impact varies based on the population structure, including the effective population size, and the rate at which recessive lethal mutations occur. In the context of livestock, the practice of intense genomic selection has led to relatively small effective population sizes (Hall

et al., 2016). With smaller effective population sizes, genetic drift can rapidly increase the frequency of recessive lethal alleles in the population. Despite the significant improvements in various traits achieved through genomic selection, such as production, fertility, and disease resistance (Gonzalez et al., 2015), this method may not offer substantial advantages in controlling the frequency of recessive lethal mutations (Dalton et al., 2015). To identify potentially deleterious alleles, a common approach is to screen populations for the absence of genomic variations in their homozygous state. If individuals in the population carry these variations in a heterozygous state, it is expected that homozygous individuals will be found based on the frequencies of these alleles and mating between carriers. This approach relies on the biological consequences of a variant's lethality, which results in the absence of the allele in its homozygous state. Detecting such lethal alleles using this approach often requires a large sample size and extensive genotyping of animals (VanRaden et al., 2011), (Pausch et al., 2015), and (Derks et al., 2017). An alternative method for identifying potentially deleterious alleles is the reverse genetic screening approach, which predicts these alleles based on the functionality of mutations. Next-generation sequencing has enabled the precise identification of potentially harmful mutations within individual genomes (Henn et al., 2015). The harmful impact of a variant can be predicted by examining its effects on gene function, including loss of function mutations such as stop-gain, stop-loss, frameshift mutations, and more (Derks et al., 2019). In this study, we analysed the dataset obtained from three Italian Pig breeds for the detection of deleterious DNA markers (SNPs and Haplotypes). We screened the dataset for genetic markers showing homozygous deficiency using two approaches single marker and haplotype-based approach. The results give a comprehensive catalog of putative deleterious alleles in the analysed pig populations aiming to improve breeding programs by excluding the deleterious genetic markers carriers.

Animals and dataset

The dataset includes three Italian pig breeds; Italian Large White (N. 5,528 samples), Italian Landrace (N. 3,470 samples), and Italian Duroc (N. 1,354 samples). The samples were analyzed using a 60K SNP-bead chip containing 57,877 markers. Markers with a call rate below 90% were removed from the dataset. In addition, only markers located on autosomal chromosomes and with a minor allele frequency greater than 0.01 were included for further analysis. Further information related to analysed samples and markers for the three analysed breeds is reported in Table1.

Single marker-based approach:

The dataset was screened for SNPs showing the absence of homozygosity. The expected homozygosity of each SNP has been computed assuming random mating by multiplying the square of minor allele frequency by the number of genotyping animals. SNPs showing the absence of homozygosity, expected to be homozygous in at least 4 samples, and showing deviation from HWE are considered to be putative deleterious SNPs. We assumed deleterious SNPs are breed-specific and a breed comparison approach was used to filter out SNPs showing the absence of homozygosity due to genotyping errors.

Haplotype based approach :

To construct haplotypes, the data was analyzed using SHAPIET2 and Ghap package in R with a window size of 400 kb and sliding windows of 200 kb, since the average extent of linkage disequilibrium in pigs is around 400kb. As recommended by (Al Bkhetan et al., 2019) we also phased the dataset using Beagle software with the same parameters. The overlapped haplotypes estimated using the two tools were considered for downstream analysis. Following (Hoff et al., 2017), when the size of marker windows increases, the probability also increases that haplotypes

of different individuals that are identical by state (IBS) are also identical by descent (IBD). We considered a threshold of 20 markers within 400kb for each haplotype as after this threshold the level of absence of homozygosity starts decreasing for the three analysed breeds. Haplotypes showing the absence of homozygosity, expected to be homozygous in at least 4 samples, and showing deviation from HWE were considered as putative deleterious haplotypes. Furthermore, the 200 kb flanked genomic region of putative deleterious markers has been annotated for candidate genes and the PigQTL databas

Genome-wide association study

We performed genome-wide association analysis for the available dataset for reproductive traits including the number of born alive, the total number of born, and the number of stillborn. An additive genetic model assuming a trend per copy of the minor allele that specifies the dependency of the number of teats on genotype categories was implemented as follows:

$$\boldsymbol{y} = \boldsymbol{W\alpha} + \boldsymbol{x}\beta + \boldsymbol{g} + \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{y}$ ($n \times 1$) is a vector containing the phenotype (the number of teats) for the n^{th} animal, $\boldsymbol{W}$ ($n \times k$) is a covariate matrix with $k = 2$ (a column of 1s and a column for the sex of the animal) and $\boldsymbol{\alpha}$ is the k-dimensional vector of covariates effects, $\boldsymbol{x}$ ($n \times 1$) is the vector containing genotypes for the i^{th} DNA marker (SNP or haplotype), β is the additive fixed effect of the i^{th} DNA marker on the phenotype, $\mathbf{g} \sim N(\mathbf{0}, \sigma^2_g \mathbf{K})$ is a multivariate Gaussian polygenic effect, with covariance matrix proportional to the relatedness matrix $\mathbf{K}$ ($n \times n$) and $\mathbf{e} \sim N(\mathbf{0}, \sigma^2_e \mathbf{I})$ is a multivariate Gaussian vector of uncorrelated residuals. Polygenic effect and residuals represented the random effects. The association between each DNA marker and the reproductive phenotype was assessed by testing the null hypothesis $H_0: \beta = 0$. Significance was tested by using the Wald test. All the models were

fitted with GEMMA v.0.98 (Zhou and Stephens 2012) after computing the relatedness matrices $\mathbf{K_1}$ and $\mathbf{K_2}$ as a centered genomic matrices, for SNPs and haplotypes, respectively. Details about the genomic matrix construction are given in the manual of GEMMA.

CNV analysis.

We used PennCNV software (Wang et al., 2007) to identify CNVs from SNP bead chip dataset of the three analysed breeds. PennCNV , incorporates LRR and BAF at each SNP marker as well as the population frequency of B allele (PFB) of SNPs into a Hidden Markov Model. Both LRR and BAF were exported from GenomeStudio for each SNP, and the PFB was generated based on the BAF of each SNP. The Illumina reports were available for 1532 samples from Italian Large White, 1010 samples from Italian Landrace, and 225 samples from the Italian Duroc breed. Single-nucleotide polymorphism physical positions on chromosomes were determined based on the Sscrofa 11 reference. Genomic waves were adjusted for the GC content of the 500 Kb genomic region surrounding each SNP on its both sides using the *-gc* model option in PennCNV. In order to decrease the number of false positive results, CNV with a length at least 50 kb and 5 SNPs within each CNV were considered.

CNVs were also identified from pooled WGS dataset for Italian Duroc and Italian Large White breeds (Bovo et al, 2020). In summary, We cleaned the reads sequences in the fastq files for the adapter removing, reads containing more than 10% unknown bases and reads containing low-quality bases ($Q \leq 5$) over 50% of the total sequenced bases were discarded. Reads were mapped on the latest version of the *Sus scrofa* reference genome (SSCROFA11.1) with BWA tool 0.7.17 (Li & Durbin, 2009; function, MEM) and the parameters for paired-end data. PICARD version 2.1.1 (https://broadinstitute.github.io/picard/) was used to remove duplicate reads. The CN.MOPS version 1.32 tool (Klambauer *et al*. 2012) was used to identify autosomal CNVs. CN.MOPS was run with

the default parameters considering the windows size of 750bp. We used bedtools to check the overlapped genomic region of putative deleterious DNA markers (SNPs or Haplotypes) with the detected CNVs.

WGS data mining :

The WGS dataset was available for 9 carriers of the putative lethal SNPs, five samples for Italian Landrace, and four for the Italian Duroc breed.. After adapter cleaning, Reads were mapped on the latest version of the *Sus scrofa* reference genome (SSCROFA11.1) with BWA tool 0.7.17 (Li & Durbin, 2009; function, MEM). Samtools was used to discard PCR duplicates. We call the variants using bcftools mpileup function (https://samtools.github.io/bcftools/bcftools.html) with parameters --consensus-caller and -q 20. We annotate the variants (SNPs, Indels, and SVs) using a Variant Effect Predictor (VEP, release 90). Using SIFT (Kumar et al., 2009) we performed The variant effect prediction in protein-altering variants. We consider the following variants as potentially causing LoF: splice acceptor, splice donor, inframe indels, frameshift, stop loss, stop gained, and start lost variants.

Results :

Putative deleterious SNPs detected by single-marker approach

The number of analysed samples and markers for the three analysed pig breeds are given in Table 1. 25 SNPs showed a complete absence of homozygosity (observed homozygosity = 0), expected to be homozygous in at least four animals and showed deviation from HWE (Figure 1). Breed comparison approach discarded those putative deleterious SNPs that have been detected in more than one analysed breed. In total, five putative deleterious SNPs were detected from the three analysed pig breeds (Table 2). In the Italian Duroc breed, Three deleterious SNPs have been detected on chr12, chr6, and chr2. The three SNPs showed an absence of homozygosity, and

deviation from Hardy-Weinberg equilibrium (P-value <0.05), and were expected to be homozygous in at least 4 samples. The carrier frequency of those three deleterious SNPs ranged from 11% to 17%. In the Italian Landrace breed, two putative deleterious SNPs have been detected on chr7 and chr9. The two SNPs were not observed in the homozygous state, showed deviation from Hardy-Weinberg equilibrium (P-value <0.05), and were expected to be homozygous in at least 4 animals. The carrier frequencies of the two deleterious SNPs for the Italian Landrace breed is about 7% for both two deleterious SNPs. Furthermore, the five deleterious SNPs identified in Italian Landrace and Italian Duroc didn't show any absence of homozygosity in the other two analysed breeds or deviation from Hardy-Weinberg equilibrium (Table 2). In Italian Large White, all putative deleterious SNPs have been discarded because they showed an absence of homozygosity or deviation from Hardy-Weinberg equilibrium in at least one of the other analysed breeds. This could be due to genotyping error for this SNP.

Putative deleterious Haplotypes detected by Haplotype-based approach

Table 3 reports the pedigree information and number of markers used for dataset phasing for the three analysed breeds. In total, nine putative deleterious haplotypes have been identified in the three analysed breeds (Table 4). The nine detected lethal haplotypes show total absence of homozygosity, deviation from Hardy-Weinberg equilibrium (P-value <0.05), expected to be homozygous in at least 4 samples, windows size of at least 20 markers in 400 kb and confirmed by both software used for haplotypes construction (SHAPIT and beagle). In the Italian Large breed, four deleterious haplotypes have been detected on chr 7, chr14, and chr5 with carrier frequency ranging from 3% to 6 % (estimated from Beagle and SHAPIET). In Italian Landrace, two deleterious haplotypes have been identified on chr1 and chr15 with carrier frequency ranging from 4% to 5% (estimated from Beagle and SHAPIET). In the Italian Duroc breed, three

deleterious haplotypes have been detected on chr3 and chr16 with frequency ranging from 1% to 6% (estimated from Beagle and SHAPIET). The top significant lethal haplotype considering the three analyzed breeds is the HAP7 detected in the Italian Duroc breed with HWE p-value 2e-06 and expected to be homozygous in at least 12 samples while not observed in the homozygous state (Table 4).

CNVs detection from SNPs dataset and WGS dataset

Using PennCNV, A total of 2,767 samples were analysed including 1532 samples from Italian Large White, 1010 samples from Italian Landrace, and 225 samples from Italian Duroc. In total, 200 CNVs regions were identified from Italian Large, 96 CNVs regions identified from Italian Landrace, and 48 CNVs regions identified from the Italian Duroc breed. The detected CNVs covered 9% of porcine genome. No overlap between the genomic regions of CNVs detected using PennCNV and the identified lethal genetic DNA markers (Haplotypes and SNPs) have been detected using bedtools intersect function. The analysis of CNVs for WGS Using CN.MOPS identified 451 CNVs for the Italian Large White breed and 1111 CNVs for the Italian Duroc breed covering 0.14% of the porcine genome. The analysis of CNVs for the WGS pooled dataset using CN.MOPS identified we identified 1111 CNVs from the Italian Duroc breed and 451 CNVs from the Italian Large White breed. The detected CNVs covered 0.28% of the reference genome in Italian Duroc and 0.14% in the Italian Large White breed. Using bedtools, no overlap was detected between the genomic regions of putative deleterious DNA markers and the identified CNVs.

Annotation of the genomic regions of putative deleterious markers

Table 5 reports the annotation of putative lethal SNPs with the closest genes and PigQtl database. In the Italian Duroc breed, the genomic region of the candidate lethal SNP on chr12 is harboring important candidate genes such as *IGF2BP1,* and *SNF8* genes. *The Insulin-Like Growth Factor 2*

mRNA Binding Protein 1 gene (*IGF2BP1)* is involved in the growth, proliferation of cells, and early embryo development. This gene has been described as some mutations within *IGF2BP1* are associated with growth traits in animals such as sheep and goats (Liu et al., 2023). *SNF8 Subunit Of ESCRT-II* gene is involved in the pathway of genes associated with low fertility in goats (Wang et al., 2021). *adhesion G protein-coupled receptor ADGRL4* is another important candidate gene located on the genomic region of the candidate lethal SNP on chr6 Knockout mice that lack *ADGRF4* show perinatal lethality in 50% of the animals (Lu et al., 2017). In the Italian Landrace breed, The genomic region of the lethal SNP located on chr7 is harboring the *Interferon Regulatory Factor 2 Binding Protein Like IRF2BPL* gene. *IRF2BPL* encodes a transcription factor that regulates neuronal networks controlling female reproductive function in nonhuman primates and rodents (Heger et al., 2017). The genomic regions of the candidate lethal SNPs are annotated to Intramuscular fat content QTL, Meat color QTL, and Number of ribs QTL according to PigQTL database. The annotation of putative lethal Haplotypes with the closest genes and PigQtl database are reported in Table 6. In Italian Large white the lethal haplotype on chr7 (HAP1) is close to the genomic region of *Dicer 1, Ribonuclease III* gene. *DICER1* is described to be associated with litter size in French Large White pigs (Chen et al., 2022).

The same genomic region is reported to explain 2.76%–9.22% of the genetic variances for sperm morphology abnormalities in pig (Zhao et al., 2020). The genomic region located on chr5 of the two putative lethal haplotypes (HAP3 and HAP4) includes candidate genes associated with reproductive traits in pigs. *Fibulin 1 (FBLN1)* gene is involved in building blood vessel wall and this gene was illustrated by a perinatal mortality of mice with homozygous knock-out phenotype (Kostka et al., 2001). In pigs, this gene is reported to be associated with the number of piglets born alive (Bergfelder et al., 2015). Another interesting gene is *EF-Hand Calcium Binding Domain 6*

EFCAB6 located on the genomic region of the deleterious haplotype (HAP4). This gene is linked to spermatogenesis by involvement in the androgen signaling pathway (Khayatzadeh et al., 2019). In the Italian Landrace breed, the genomic region of deleterious haplotype HAP5 includes *Regulator Of G Protein Signaling 17*(*RGS17*) gene. This gene is reported to be associated with reproductive performance in cattles (Sweett et al., 2020). In Italian Duroc, the genomic region of deleterious haplotype (HAP9) is harboring the *autism susceptibility candidate 2 (AUTS2)* gene. This gene is associated with reproductive traits in pigs (sato et al., 2016). The detected deleterious haplotypes are annotated with important QTLs associated with number of born alive, the number of mummified piglets, sperm mortality and Litter size according to PigQTL database.

WGS data mining for identified deleterious markers.

Table 7 reports the Loss of function markers detected from the 9 WGS dataset for the carriers of genomic region-identified lethal SNPs. The loss of function identified markers include a frameshift variant, Start lost, Splice donor variant, and Splice acceptor variant. Those loss of function are close to a number of candidate genes (*IGF2BP1, ADGRL4, POMT2, and HGF)* that have been reported to be associated with litter size, perinatal lethality in mice, male neonatal death, and embryonically lethal in mice.

GWAS analysis for reproductive traits.

The reproductive phenotypes were available for 1558 samples in the ILA breed including the number of born alive piglets, the number of stillborn piglets, and the total number of born piglets. Table 8. Reports the information for a dataset for GWAS analysis single marker approach in ILA

breed. We ran a GWAS analysis to check the association of putative deleterious SNPs on the analysed reproductive traits. The putative deleterious SNP on chr9: 98059641 with p-value =0.02 for a number of stillborn traits. Table 9 reports the GWAS results related to putative deleterious SNPs in the Italian Landrace breed. The dataset of previous GWAS analysis using a haplotype-based approach was available for ILW breed. Table 10. Reports the information for a dataset for GWAS analysis Haplotype-based approach in ILW breed. The top significant deleterious haplotype is HAP3 with P-value = 0,001.

Table 11. reports the full results of GWAS analysis related to putative deleterious haplotypes in ILW breed.

Discussion

This study reports 14 putative lethal DNA markers including five putative lethal SNPs and nine lethal haplotypes. We show that large-scale genotype data analysis using a single marker approach and haplotype-based approach provides a powerful tool to investigate deleterious alleles. We also assume that deleterious alleles are breed-specific alleles. Based on this assumption we discarded lethal alleles identified from more than one analysed breed. The deleterious alleles detected from more than one breed could be false positive results due to genotyping errors. The carrier frequencies range from 7% -17% for SNPs and 3-6% for haplotypes in the three pig populations. This proves that the haplotype-based approach is more effective in identifying deleterious alleles with lower frequency that could not be tagged by the single-marker approach (Howard et al., 2017). In the haplotype-based approach, we used two different tools for haplotype construction (Beagle and SHAPIET2). As suggested by (Al Bkhetan et al., 2019), some haplotypes constructed by different tools are false positives and the best estimation is overlapped haplotypes estimated using more than one tool. The CNVs detected from the SNP bead chip dataset and WGS dataset do not

overlap with the genomic regions of putative lethal SNPs and Haplotypes. This confirmed that the absence of homozygous individuals for those genomic regions is not due to the presence of CNVs within these genomic regions. It is worth mentioning that the genomic regions of the putative deleterious DNA markers (SNPs and Haplotypes) identified using a single marker and haplotype-based approaches were not overlapped. WGS data information from the carriers of putative lethal haplotypes and SNPs is still needed here to confirm the position of the lethal variant causing the homozygous deficiency. According to GWAS analysis for reproductive traits, the most significant putative deleterious DNA marker is HAP3 with p-value=0.001 for the total number of born traits. The beta value for this haplotype is -1.453 which means that this haplotype is associated with decreasing the number of born alive piglets in the Italian Large White breed. Interestingly, Two of the deleterious haplotypes detected in the Italian Large White breed are located in one genomic region located *about* 1Mb (HAP3 and HAP4 Table 4). The absence of homozygosity for those two haplotypes could be due to a loss of function variant affecting the genomic region of the two haplotypes. Anaysing nine samples of the carriers of putative deleterious SNPs using the WGS dataset shows that the genomic regions of the carriers of putative deleterious SNPs include loss of function variants that affect important candidate genes such as *IGF2BP1* and *ADGRL4. The Insulin-Like Growth actor 2 mRNA Binding Protein 1* gene (*IGF2BP1)* is involved in the growth, proliferation of cells, and early embryo development (Wu et al., 2020). *adhesion G protein-coupled receptor ADGRL4* gene shows perinatal lethality in mice (Lu et al., 2017). We were not able to screen the genomic region of putative lethal Haplotypes for loss of function variants due to the absence of the WGS dataset of the carriers of those lethal DNA markers. In the future, we still need to sequence the carriers of the putative lethal SNPs and Haplotypes to identify the lethal variants within the genomic regions of those deleterious DNA markers. Larger sample sizes

including more amount of phenotypic records and genotype datasets collected from the three anlysed pig breeds could improve the detection and management of the deleterious alleles with low frequency.

References:

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, *25*(14), 1754-1760.

Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D. A., Mitterecker, A., Bodenhofer, U., & Hochreiter, S. (2012). cn. MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic acids research*, *40*(9), e69-e69.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7

Al Bkhetan, Z., Zobel, J., Kowalczyk, A., Verspoor, K., & Goudey, B. (2019). Exploring effective approaches for haplotype block phasing. *BMC bioinformatics*, *20*(1), 1-14.

Bergfelder-Drüing, S., Grosse-Brinkhaus, C., Lind, B., Erbe, M., Schellander, K., Simianer, H., & Tholen, E. (2015). A genome-wide association study in large white and landrace pig populations for number piglets born alive. PloS one, 10(3), e0117468.

Bovo, S., Ribani, A., Munoz, M., Alves, E., Araujo, J. P., Bozzi, R., ... & Fontanesi, L. (2020). Genome-wide detection of copy number variants in European autochthonous and

commercial pig breeds by whole-genome sequencing of DNA pools identified breed-characterising copy number states. *Animal Genetics*, *51*(4), 541-556.

Bovo, S., Ribani, A., Munoz, M., Alves, E., Araujo, J. P., Bozzi, R., ... & Fontanesi, L. (2020). Genome-wide detection of copy number variants in European autochthonous and commercial pig breeds by whole-genome sequencing of DNA pools identified breed-characterising copy number states. *Animal Genetics*, *51*(4), 541-556.

Chen, J., Wu, Z., Chen, R., Huang, Z., Han, X., Qiao, R., ... & Li, X. L. (2022). Identification of genomic regions and candidate genes for litter traits in French large white pigs using genome-wide association studies. Animals, 12(12), 1584.

El-Dessouky, S. H., Hosny, H., Elarab, A. E., & Issa, M. Y. (2020). Prenatal presentation of Walker–Warburg syndrome with a POMT2 mutation: an extended fetal phenotype. *Egyptian Journal of Medical Human Genetics*, *21*, 1-6.

Heger S, Mastronardi C, Dissen GA, Lomniczi A, Cabrera R, Roth CL, et al. Enhanced at puberty 1 (EAP1) is a new transcriptional regulator of the female neuroendocrine reproductive axis. J Clin Invest. 2007;117.

Howard, D. M., Pong-Wong, R., Knap, P. W., & Woolliams, J. A. (2017). Use of haplotypes to identify regions harbouring lethal recessive variants in pigs. *Genetics Selection Evolution*, *49*(1), 1-10

Hoff, Jesse L., et al. "Lethal Haplotypes and Candidate Causal Mutations in Angus Cattle." bioRxiv (2017): 125203.

Khayatzadeh, N., Mészáros, G., Utsunomiya, Y. T., Schmitz-Hsu, F., Seefried, F., Schnyder, U., ... & Sölkner, J. (2019). Genome-wide mapping of the dominance effects based on breed

ancestry for semen traits in admixed Swiss Fleckvieh bulls. Journal of dairy science, 102(12), 11217-11224.

Kostka G, Giltay R, Bloch W, Addicks K, Timpl R, et al. (2001) Perinatal lethality and endothelial cell abnormalities in several vessel compartments of fibulin-1-deficient mice. Mol Cell Biol 21: 7025–7034. pmid:11564885

Liu, H., Li, H., Mao, C., Jiang, F., Hao, K., Han, X., ... & Lan, X. (2021). Insights into genetic variants within sheep IGF2BP1 and their association with litter size. *Small Ruminant Research*, *198*, 106350.

Liu, H., Xu, H., Lan, X., Cao, X., & Pan, C. (2023). The InDel variants of sheep IGF2BP1 gene are associated with growth traits. Animal Biotechnology, 34(2), 134-142.

Lu S, Liu S, Wietelmann A, Kojonazarov B, Atzberger A, Tang C, et al. Developmental vascular remodeling defects and postnatal kidney failure in mice lacking Gpr116 (Adgrf5) and Eltd1 (Adgrl4). PLoS One. 2017;12:e0183166.

Lu, S., Liu, S., Wietelmann, A., Kojonazarov, B., Atzberger, A., Tang, C., ... & Offermanns, S. (2017). Developmental vascular remodeling defects and postnatal kidney failure in mice lacking Gpr116 (Adgrf5) and Eltd1 (Adgrl4). *PLoS One*, *12*(8), e0183166.

Sato, S., Kikuchi, T., Uemoto, Y., Mikawa, S., & Suzuki, K. (2016). Effect of candidate gene polymorphisms on reproductive traits in a Large White pig population. Animal Science Journal, 87(12), 1455-1463.

Shao, X., Tan, C., Chen, D., & Wang, Y. L. (2018). Placental defects are involved in most gene mutations that cause embryonic and fetal death. *Biology of Reproduction*, *99*(3), 476-478.

Sweett, H., Fonseca, P. A. S., Suárez-Vega, A., Livernois, A., Miglior, F., & Cánovas, A. (2020). Genome-wide association study to identify genomic regions and positional candidate genes associated with male fertility in beef cattle. Scientific reports, 10(1), 20102.

Wang, K., Liu, X., Qi, T., Hui, Y., Yan, H., Qu, L., ... & Pan, C. (2021). Whole-genome sequencing to identify candidate genes for litter size and to uncover the variant function in goats (Capra hircus). Genomics, 113(1), 142-150.

Wu, J., Lu, C., Ge, S., Mei, J., Li, X., & Guo, W. (2020). Igf2bp1 is required for hepatic outgrowth during early liver development in zebrafish. *Gene, 744, 144632.*

Zhao, Y., Gao, N., Li, X., El-Ashram, S., Wang, Z., Zhu, L., ... & Li, Z. (2020). Identifying candidate genes associated with sperm morphology abnormalities using weighted single-step GWAS in a Duroc boar population. Theriogenology, 141, 9-15.

Tables

Table 1. The number of analysed samples and markers for the three analysed pig breeds are given

Breed	The number of analysed samples	Number of analysed SNPs	The number of analysed samples (filtered dataset)	Number of analysed SNPs (filtered dataset)
Italian Large White	5,528	57,877	5,368	50,929
Italian Landrace White	3,470	57,877	3,470	50,435
Italian Duroc White	1,354	57,877	1,354	45,153

Table 2. Putative deleterious SNPs detected in the analysed pig breeds.

SNP	Breed	position	Genotype	Genotype for other analysed breeds	HWE-Pvalue	Expected homozygosity	Carrier frequency
ALGA0119814	DU	chr12:25319584	0/230/1120	LA:(656/1695/984), LW:(836/2394/1898)	0,000034	9,796	0,170
WU_10.2_6_124428986	DU	chr6:134186379	0/151/1203	LA:(859/1712/895), LW:(1280/2680/1404)	0,0188	4,210	0,112
ASGA0010561	DU	chr2:82318055	0/150/1202	LA:(203/1110/2153), LW:(931/2477/1955)	0,0313	4,161	0,111
M1GA0010684	LA	chr7:100084753	0/266/3204	DU:(78/484/787), LW:(441/2090/2829)	0,0094	5,098	0,077
ALGA0054305	LA	chr9:98059641	0/246/3223	LA:(103/575/673), LW:(103/575/673)	0,0218	4,361	0,071

Table 3. Presents comprehensive details on haplotypes estimated using SHAPEIT2 and Beagle across three analysed breeds.

Breed	N.samples	markers	Filtered samples	Filtered markers	DIOS	TRIOS	Haplotypes SHAPIT2	Haplotype Beagle	Overlapped
LW	5,528	57,877	5,368	50,929	2,162	347	78,561	72,116	87%
LA	3,47	57,877	3,47	50,435	1,221	438	81,996	64,517	75%
DU	1,354	57,877	1,354	45,153	668	0	67,83	51,636	71%

Table 4. Putative deleterious Haplotypes were detected in the three analysed breeds.

Candidates haplotype	CHR	POS1	Breed	POS2	N.carrier	carrier-freq	MAF	OBS-HOM	N.markers	HWE	Expected-homozygous
HAP1	7	116200001	LW	116600000	421	3%	0,040	0	20	2e-06	8,511
HAP2	14	1400001	LW	1800000	345	6%	0,033	0	26	0,003294	5,716
HAP3	5	4000001	LW	4400000	331	5%	0,032	0	21	0,005189	5,261
HAP4	5	4600001	LW	5,00E+06	320	5%	0,031	0	27	0,007318	4,917
HAP5	1	12800001	LA	13200000	302	5%	0,045	0	24	0,001406	6,567
HAP6	15	1200001	LA	1600000	288	4%	0,043	0	20	0,002548	5,972
HAP7	3	2200001	DU	2600000	264	6%	0,097	0	23	2,63E-06	12,850
HAP8	16	64400001	DU	64800000	198	2%	0,073	0	21	0,000726	7,228
HAP9	3	14200001	DU	14600000	186	1%	0,069	0	24	0,001698	6,378

Table 5. Loss of function markers detected from WGS dataset for the carriers of genomic region identified lethal SNPs

SNP	Breed	position	Annotated genes	PigQtl database
ALGA0119814	DU	chr12:2531958 4	*UBE2Z#**SNF8**#**IGF2BP1**#B4GALNT2#PHOSPHO1#ZNF652#GIP*	Conformation score QTL (126148)
WU_10.2_6_12442898 6	DU	chr6:13418637 9	***ADGRL4***	Backfat at tenth rib QTL (17979) Meat color QTL (283230) Mean corpuscular volume_QTL (107473)#Subcutaneous_fat_thickness_QTL_(283225)
ASGA0010561	DU	chr2:82318055	*TMEM174#TMEM171#FCHO2*	
M1GA0010684	LA	chr7:10008475 3	*MIR9860#VASH1#ANGEL1#LRRC74A#**IRF2BPL**#CIPC#ZDHHC22* *TMEM63C*	Number of ribs QTL (275341)
ALGA0054305	LA	chr9:98059641	*CACNA2D1*	Intramuscular fat content QTL_(125483)

Table 6. The annotation for the genomic regions of candidates' lethal Haplotypes with the closest genes and PigQTL database.

Candidates haplotype	CHR	Annotated genes+-200kb	PigQTL	Breed
HAP1	7	*GSC#GLRX5# SYNE3#CLMN#**DICER1***	Litter size QTL_(261987), Sperm motility (QTL:160673)	ILW
HAP2	14	*DIRAS2*	NA	ILW
HAP3	5	*RIBC2#**FBLN1**#FAM118A#UPK3A #NUP50#ARHGAP8#PRR5# SMC1B#PHF21B#KIAA0930*	Litter size QTL (64739)	ILW
HAP4	5	*ARHGAP8#PRR5#RTL6#KIAA1644 #PARVG#SAMM50#PNPLA3#SULT4A1#PARVB#EFCAB6*	Intramuscular fat content QTL (176605) #Meat color QTL (55953)	ILW
HAP5	1	*OPRM1#RGS17#IPCEF1*	NA	ILA
HAP6	15	*RND3*	Birth_weight variability QTL (95468)4th last_ribs_QTL (139260)#Number_of_mummified_pigs QTL (178866)	ILA
HAP7	3	*CARD11#GNA12*	Litter weight QTL (281339)	DU
HAP8	16	*UBLCP1#RNF145#EBF1*	NA	DU
HAP9	3	*AUTS2*	Total number born alive QTL (281338)#Litter size_QTL_(130309)	DU

Table 7. The annotation for the genomic regions of candidates' lethal Haplotypes with the closest genes and PigQTL database.

Genomic position	Breed	Type	Ref-ALT	Candidate-Gene	Phenotype	Reference	Distance form lethal SNP
12_25152745	DU	Frameshift variant	C/T	*IGF2BP1*	litter size	(Liu et al. 2021)	166
6_133953269	DU	Start lost	A/G	*ADGRL4*	perinatal lethality in mice	(Lu et al. 2017)	233
7_100541270	ILA	Splice donor variant	C/A	*POMT2*	male neonatal death	(El-Dessouky et al. 2020)	456
9_98500535	ILA	Splice acceptor variant	T/G	*HGF*	Embryonically lethal in mice	Shao, Xuan et al .2018	440

Table 8. Information for a dataset for GWAS analysis single marker approach in ILA breed.

Information	N. Born alive	Fraction of stillborn to born alive	N. stillborn	Total number of born
number of total individuals	1558	1558	1558	1558
number of analyzed individuals	1558	1558	1558	1558
number of covariates	1	1	1	1
number of phenotypes	1	1	1	1
number of total SNPs/var	59509	59509	59509	59509
number of analyzed SNPs/var	59509	59509	59509	59509
PVE	0,079	0,011	0,007	0,069
se(PVE)	0,028	0,021	0,022	0,028

Table 9. Reports the GWAS results related to putative deleterious SNPs in the Italian Landrace breed.

Parity	TRAIT	SNP	pos	a1	a2	af	beta	se	logl_H1	l_remle	p_wald
P4	FRACTION	ALGA0054305	98059641	A	G	0,038	2,08E+00	9,18E-01	-3,71E+03	1,11E-02	0,024
P4	SB	ALGA0054305	98059641	A	G	0,038	-2,55E-01	1,14E-01	-1,49E+03	9,98E-03	0,026
P2	SB	ALGA0054305	98059641	A	G	0,036	-2,22E-01	1,07E-01	-2,17E+03	1,00E-05	0,037
P2	FRACTION	ALGA0054305	98059641	A	G	0,036	1,76E+00	8,45E-01	-5,21E+03	1,00E-05	0,037
P3	SB	ALGA0054305	98059641	A	G	0,037	-2,25E-01	1,15E-01	-1,92E+03	9,33E-02	0,050
P3	SB	M1GA0010684	100084753	A	G	0,034	2,20E-01	1,19E-01	-1,92E+03	6,96E-02	0,064
P3	FRACTION	ALGA0054305	98059641	A	G	0,037	1,60E+00	9,20E-01	-4,66E+03	3,86E-02	0,083
P3	FRACTION	M1GA0010684	100084753	A	G	0,034	-1,57E+00	9,55E-01	-4,66E+03	2,03E-02	0,100
P4	SB	M1GA0010684	100084753	A	G	0,033	1,66E-01	1,22E-01	-1,49E+03	1,96E-04	0,175
P5	FRACTION	M1GA0010684	100084753	A	G	0,031	1,45E+00	1,10E+00	-2,77E+03	8,81E-02	0,187
P4	FRACTION	M1GA0010684	100084753	A	G	0,033	-1,26E+00	9,80E-01	-3,71E+03	2,73E-03	0,199
P5	SB	M1GA0010684	100084753	A	G	0,031	-1,85E-01	1,47E-01	-1,14E+03	1,40E-01	0,206

P4	BA	M1GA0010684	100084753	A	G	0,033	-4,18E-01	3,60E-01	-2,64E+03	3,94E-02	0,247
P1	TN	M1GA0010684	100084753	A	G	0,034	-2,90E-01	2,59E-01	-3,61E+03	2,00E-01	0,262
P4	BA	ALGA0054305	98059641	A	G	0,038	3,55E-01	3,38E-01	-2,64E+03	3,98E-02	0,294
P1	BA	M1GA0010684	100084753	A	G	0,034	-2,81E-01	2,70E-01	-3,66E+03	2,32E-01	0,298
P5	BA	ALGA0054305	98059641	A	G	0,037	3,61E-01	3,62E-01	-1,95E+03	1,00E-05	0,319
P5	TN	ALGA0054305	98059641	A	G	0,037	3,67E-01	3,70E-01	-1,97E+03	1,00E-05	0,321
P3	TN	M1GA0010684	100084753	A	G	0,034	2,83E-01	3,09E-01	-3,18E+03	2,77E-02	0,360
P2	BA	ALGA0054305	98059641	A	G	0,036	2,54E-01	3,01E-01	-3,65E+03	8,70E-02	0,399
P1	BA	ALGA0054305	98059641	A	G	0,036	2,13E-01	2,66E-01	-3,67E+03	2,29E-01	0,423
P1	FRACTION	ALGA0054305	98059641	A	G	0,036	6,01E-01	8,20E-01	-5,49E+03	3,10E-02	0,464
P3	BA	ALGA0054305	98059641	A	G	0,037	2,20E-01	3,01E-01	-3,21E+03	2,15E-02	0,465
P4	TN	M1GA0010684	100084753	A	G	0,033	-2,49E-01	3,57E-01	-2,62E+03	6,96E-02	0,485
P1	SB	ALGA0054305	98059641	A	G	0,036	-6,31E-02	9,89E-02	-2,20E+03	2,13E-02	0,523
P1	TN	ALGA0054305	98059641	A	G	0,036	1,42E-01	2,55E-01	-3,61E+03	2,00E-01	0,578
P2	TN	M1GA0010684	100084753	A	G	0,033	1,25E-01	3,07E-01	-3,64E+03	9,41E-02	0,684
P2	SB	M1GA0010684	100084753	A	G	0,033	4,40E-02	1,10E-01	-2,17E+03	1,00E-05	0,690

P5	BA	M1GA0010684	100084753	A	G	0,031	1,52E-01	3,95E-01	-1,95E+03	1,00E-05	0,700
P1	SB	M1GA0010684	100084753	A	G	0,034	-3,05E-02	1,01E-01	-2,20E+03	1,70E-02	0,762
P2	BA	M1GA0010684	100084753	A	G	0,033	8,11E-02	3,10E-01	-3,65E+03	9,29E-02	0,794
P4	TN	ALGA0054305	98059641	A	G	0,038	7,78E-02	3,35E-01	-2,62E+03	6,94E-02	0,816
P2	FRACTION	M1GA0010684	100084753	A	G	0,033	-1,94E-01	8,73E-01	-5,22E+03	1,00E-05	0,824
P3	BA	M1GA0010684	100084753	A	G	0,034	6,25E-02	3,15E-01	-3,21E+03	2,07E-02	0,843
P5	FRACTION	ALGA0054305	98059641	A	G	0,037	1,07E-01	1,02E+00	-2,77E+03	7,96E-02	0,916
P5	SB	ALGA0054305	98059641	A	G	0,037	1,15E-02	1,37E-01	-1,14E+03	1,27E-01	0,933
P1	FRACTION	M1GA0010684	100084753	A	G	0,034	4,44E-02	8,36E-01	-5,49E+03	2,95E-02	0,958
P2	TN	ALGA0054305	98059641	A	G	0,036	1,28E-02	2,99E-01	-3,64E+03	9,52E-02	0,966
P5	TN	M1GA0010684	100084753	A	G	0,031	-1,36E-02	4,03E-01	-1,97E+03	1,00E-05	0,973
P3	TN	ALGA0054305	98059641	A	G	0,037	5,95E-03	2,98E-01	-3,18E+03	3,85E-02	0,984

Table 10. Information for a dataset for GWAS analysis single marker approach in ILW breed.

Information	N. Born alive	Fraction of stillborn to born alive	N. stillborn	Total number of born
number of total individuals	1395	1395	1395	1395
number of analyzed individuals	1395	1395	1395	1395
number of covariates	1	1	1	1
number of phenotypes	1	1	1	1
number of total SNPs/var	246415	246415	246415	246415
number of analyzed SNPs/var	246415	246415	246415	246415
PVE	0,000	0,000	0,000	0,009
se(PVE)	0,030	0,067	0,081	0,028

Table 11. Reports the full results of GWAS analysis related to putative deleterious haplotypes in ILW breed

HAP	chr	st	end	Haploblock	AF	p_wald	Parity	trait
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,0019	P2	Total number of born
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,0033	P2	BA
HAP2	14	1200001	2000001	CHR14_B8	0,036	0,0054	P5	Number of stillborn
HAP2	14	1200001	2000001	CHR14_B8	0,028	0,0115	P1	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,036	0,0162	P5	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,028	0,0363	P1	Number of stillborn
HAP2	14	1200001	2000001	CHR14_B8	0,036	0,0426	P5	BA
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,023	0,0582	P3	BA
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,1423	P4	Total number of born
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,1813	P3	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,1843	P3	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,023	0,1891	P3	Total number of born
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,1924	P4	BA
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,023	0,2651	P3	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,023	0,2735	P3	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,3187	P2	Fraction Born death/born alive
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,3482	P1	BA
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,3968	P1	Total number of born
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,4508	P2	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,4544	P2	Total number of born

HAP3	5	4400001	5200001	CHR5_B24	0,011	0,4863	P5	BA
HAP2	14	1200001	2000001	CHR14_B8	0,028	0,5049	P1	BA
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,5237	P2	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,5265	P3	BA
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,5502	P5	Total number of born
HAP2	14	1200001	2000001	CHR14_B8	0,036	0,5528	P5	Total number of born
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,5572	P5	Fraction Born death/born alive
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,5647	P1	BA
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,024	0,5683	P4	Total number of born
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,5717	P3	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,5868	P1	Fraction Born death/born alive
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,6072	P3	Total number of born
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,6304	P1	Number of stillborn
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,6379	P2	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,018	0,6685	P5	Number of stillborn
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,6731	P1	Fraction Born death/born alive
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,6810	P4	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,6820	P2	BA
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,7013	P3	Fraction Born death/born alive
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,024	0,7047	P4	BA
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,024	0,7290	P4	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,024	0,7376	P4	Fraction Born death/born alive
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,7554	P3	BA
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,7642	P2	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,018	0,7716	P5	Fraction Born death/born alive
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,7741	P1	Total number of born
HAP2	14	1200001	2000001	CHR14_B8	0,028	0,7741	P1	Total number of born
HAP3	5	4400001	5200001	CHR5_B24	0,011	0,7927	P5	Number of stillborn
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,022	0,8046	P1	Number of stillborn
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,8150	P2	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,032	0,8268	P4	Fraction Born death/born alive

HAP2	14	1200001	2000001	CHR14_B8	0,032	0,8565	P4	Total number of born
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,018	0,8571	P5	Total number of born
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,8760	P2	BA
HAP3	5	4400001	5200001	CHR5_B24	0,012	0,9134	P4	Fraction Born death/born alive
HAP2	14	1200001	2000001	CHR14_B8	0,032	0,9203	P4	BA
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,9341	P3	Total number of born
HAP1	7	1,16E+08	1,17E+08	CHR7_B582	0,018	0,9577	P5	BA
HAP2	14	1200001	2000001	CHR14_B8	0,029	0,9789	P2	Total number of born
HAP2	14	1200001	2000001	CHR14_B8	0,032	0,9917	P4	Number of stillborn

Figures

Figure1. The genomic distribution of the putative deleterious DNA markers.

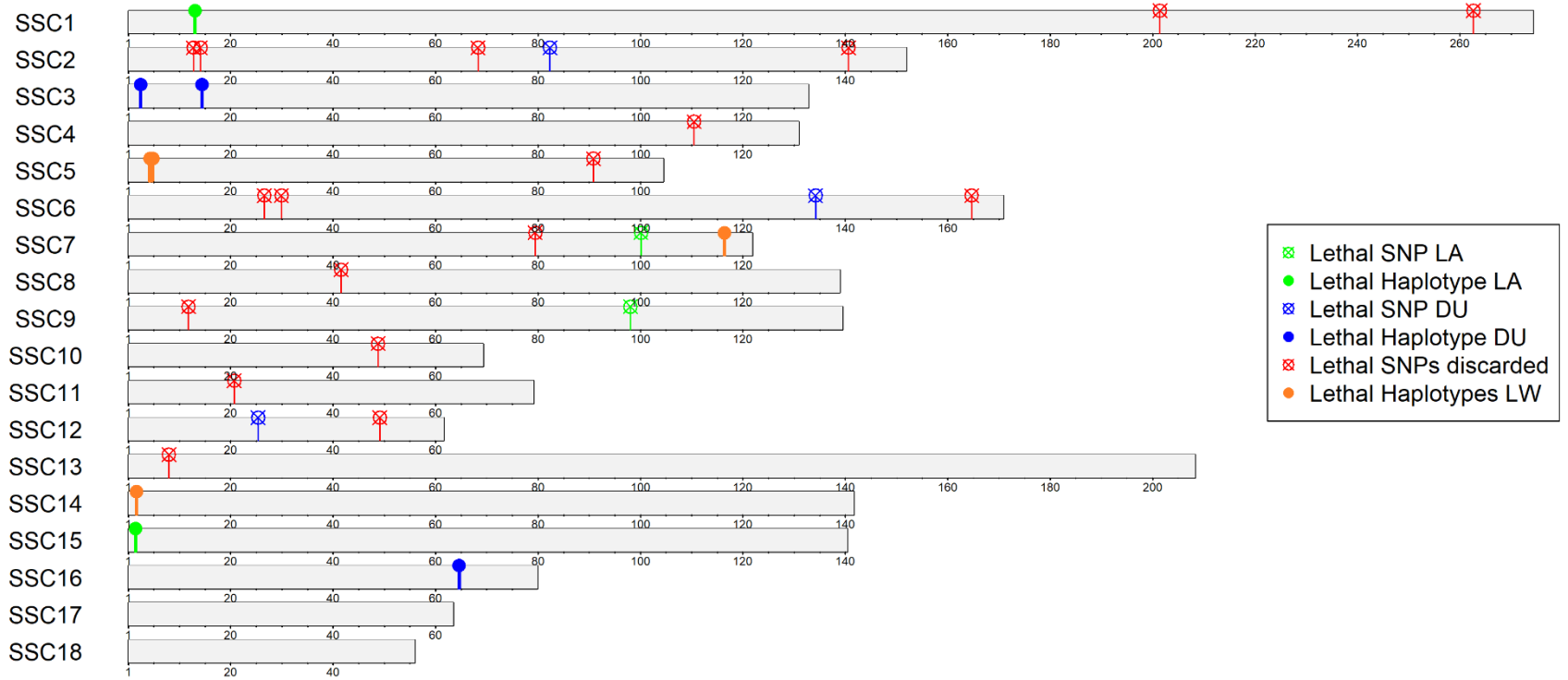
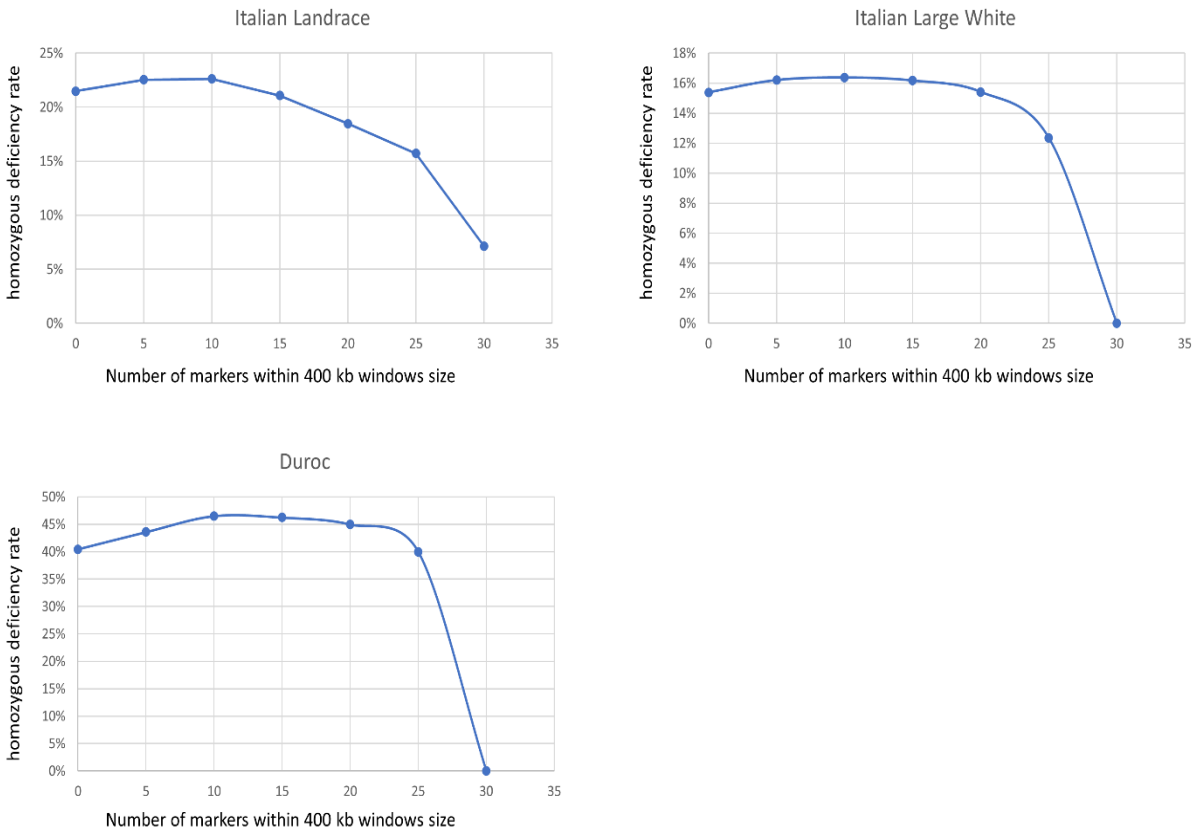


Figure2. The level of homozygous deficiency with different numbers of markers within each 400kb window size.



Chapter 3: **Genomic diversity and signatures of selection in meat and fancy rabbit breeds based on high-density marker data.**

Mohamad Ballan[1§], Samuele Bovo[1§], Giuseppina Schiavo[1], Michele Schiavitto[2], Riccardo Negrini[3] and Luca Fontanesi[1*]

[1]Division of Animal Sciences, Department of Agricultural and Food Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy

[2]Associazione Nazionale Coniglicoltori Italiani (ANCI), Contrada Giancola snc, 71030 Volturara Appula (FG), Italy

[3]Associazione Italiana Allevatori, Via G. Tomassetti 9, 00161 Roma, Italy

*Corresponding author

[§]Equal contribution

Published in Genetics Selection Evolution on 21 January 2022. doi: https://doi.org/10.1186/s12711-022-00696-9

Abstract

Background

Domestication of the rabbit (*Oryctolagus cuniculus*) has led to a multi-purpose species that includes many breeds and lines with a broad phenotypic diversity, mainly for external traits (e.g. coat colours and patterns, fur structure, and morphometric traits) that are valued by fancy rabbit breeders. As a consequence of this human-driven selection, distinct signatures are expected to be present in the rabbit genome, defined as signatures of selection or selective sweeps. Here, we investigated the genome of three Italian commercial meat rabbit breeds (Italian Silver, Italian Spotted and Italian White) and 12 fancy rabbit breeds (Belgian Hare, Burgundy Fawn, Champagne d'Argent, Checkered Giant, Coloured Dwarf, Dwarf Lop, Ermine, Giant Grey, Giant White, Rex, Rhinelander and Thuringian) by using high-density single nucleotide polymorphism data. Signatures of selection were identified based on the fixation index (F_{ST}) statistic with different approaches, including single-breed and group-based methods, the latter comparing breeds that are grouped based on external traits (different coat colours and body sizes) and types (i.e. meat vs. fancy breeds).

Results

We identified 309 genomic regions that contained signatures of selection and that included genes that are known to affect coat colour (*ASIP*, *MC1R* and *TYR*), coat structure c and body size (*LCORL*/*NCAPG*, *COL11A1* and *HOXD*) in rabbits and that characterize the investigated breeds. Their identification proves the suitability of the applied methodologies for capturing recent selection events. Other regions included novel candidate genes that might contribute to the phenotypic variation among the analyzed breeds, including genes for pigmentation-related traits (*EDNRA*, *EDNRB*, *MITF* and *OCA2*) and body size, with a strong candidate for dwarfism in rabbit (*COL2A1*).

Conclusions

We report a genome-wide view of genetic loci that underlie the main phenotypic differences in the analyzed rabbit breeds, which can be useful to understand the shift from the domestication process to the development of breeds in *O. cuniculus*. These results enhance our knowledge about the major genetic loci involved in rabbit external traits and add novel information to understand the complexity of the genetic architecture underlying body size in mammals.

Background

The European rabbit (*Oryctolagus cuniculus*), usually simply referred to as rabbit, is the only species that has been domesticated exclusively in western Europe. Its domestication started from the wild populations in the South of France that originally derived from the wild populations of the *O. c. cuniculus* subspecies spread in the Iberian Peninsula, which experienced a postglacial expansion (reviewed in [1]). Among the possible animal domestication trajectories that have been theorised [2], domestication of the rabbit better matches the directed pathway that does not involve preliminary steps of habituation of animals to human beings and begins with the capture of wild animals, with the aim of controling their breeding and reproduction. It seems plausible that this process occurred quite recently in rabbit, starting in the French monasteries and castles in the High Middle Ages and, continuing until the XV-XVI centuries [3, 4], and then may have continued through the dispersion and transfer of rabbits in the North of Europe, until the most recent constitution of some modern breeds [5]. Domestication of the rabbit occurred after a first genetic bottleneck that involved the wild subpopulations from which the domestic lines were then derived, accompanied by limited recurrent introgression from the wild types [6, 7]. However, this resulted in only slightly modified allele frequencies at many loci between the wild and domestic rabbit populations, which suggests that the domestication process had a relatively weak effect on standing genetic variation in many regulatory regions of the genome [6]. These changes occurred mainly in

genes that are involved in brain and neuronal development, which indicates that the resulting modified behaviour traits were important for the domestic rabbit to adapt to the human environment. Thus, in this species, the domestication process relied on derived genetic material, which contained variants that determine favourable behavioral traits and facilitate handling and breeding [6].

The domestication process was integrated or was followed by artificial selection processes that led to the constitution of many breeds. The resulting rabbit breeds can be distinguished based on their broad phenotypic diversity, in terms of external traits that are valued by fancy breeders [1, 8, 9]. It is also worth mentioning that the most recently constituted rabbit lines or strains and the modern breeds have been derived by cross-breeding pre-existing varieties or morphs, with the aim to introgress desirable traits to create new lines or combinations of morphological features or to improve production traits in specialized meat lines [1, 8, 9]. However, coat colours and coat colour patterns are the most relevant traits that differentiate many rabbit breeds, as also demonstrated by the fact that many breeds are named according to their colouration [1, 9].

In rabbits as well as in many other mammals, several coat colour loci were first described by classical genetic studies that confirmed the Mendelian segregation of these inherited traits and established homology across species [10]. Subsequent molecular characterizations identified the causal mutations or associated markers at some of these loci. At the *extension* locus, three mutated alleles (E^D or E^S, determining the dominant black/steel coat colour; e^J, determining the Japanese brindling pattern identified in the Japanese and Rhinelander breeds [12]; *e*, determining the recessive yellow/red coat colour) are caused by mutations in the *melanocortin 1 receptor* (*MC1R*) gene [11, 12]. The *agouti* locus is determined by mutations in the *agouti signaling protein* (*ASIP*) gene that form recessive the black non-agouti (*a*) and tan (a^t) alleles [13, 14]. Several alleles at the *albino* locus (*C* series) are caused by mutations in the *tyrosinase* (*TYR*) gene that produce the

chinchilla, the Himalayan (of the Californian breed), and the full albino coat colours (of the New Zealand White and related breeds and populations) [15, 16]. At the *English spotting* locus, a marker in the *v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog* (*KIT*) gene is associated with the spotted pattern of the Checkered Giant and Rhinelander breeds, whose classical spotted design is due to the heterozygous genotype *En/en* that preserves these animals from a megacolon defect associated with the *En* allele [17]. Other coat colour loci might be involved in the spotted phenotypes of rabbits but, to date, the corresponding genes have not been identified [1, 10].

In addition to coat colour, breeds can be distinguished by their hair structure. For example, a mutation in the *lipase member H* (*LIPH*) gene determines the *Rex* locus R^1 [18] that confers soft down-hair. The shape and position of the ears are other morphological traits that differentiate some breeds (e.g. several lop breeds). Another main morphological feature of rabbit breeds is body size: breeds are traditionally classified into dwarf, small, medium, and large classes according to their adult body weight [9]. Carneiro *et al.* [19] identified a large deletion in the *high mobility group AT-hook 2* (*HMGA2*) gene as the causal mutation for one type of dwarfism in rabbit. However, other loci might also contribute to the reduced size of some dwarf rabbits and several loci involved in dwarfism have been reported in this species (reviewed in [20]).

Although several studies have successfully started to dissect the genetic mechanisms underlying some external traits in rabbit, these mainly focused on a few candidate genes, and a complete genetic picture of the phenotypic diversity of many rabbit breeds is still lacking, not only in terms of coat colour but also in terms of body size, meat production and performance traits. Few studies in this species have been designed to investigate the variability at the genome-wide level and identify footprints of recent selection that distinguish rabbit breeds [6, 19, 21].

In the current study, we used high-density single nucleotide polymorphism (SNP) genotyping data from three commercial meat breeds and 12 fancy breeds that differ for several external traits to

evaluate the level of genetic diversity and identify signatures of selection in the rabbit genome that may explain the phenotypic variability that differentiates these breeds.

Methods

Animals

Biological specimens (hair roots or buccal swaps) were collected from 660 rabbits from 15 breeds, including three commercial meat lines (Italian Silver, Italian Spotted and Italian White) and 12 fancy breeds (Belgian Hare, Burgundy Fawn, Champagne d'Argent, Checkered Giant, Coloured Dwarf, Dwarf Lop, Ermine, Giant Grey, Giant White, Rex, Rhinelander and Thuringian). All rabbits had the standard breed characteristics, as registered in the corresponding breed herd book maintained by the Italian Rabbit Breeders Association (ANCI). The description of the breeds and the number of animals analysed for each breed are in Table 1. Sampled rabbits were chosen to avoid highly related individuals (no full- or half-sibs).

Genotyping

DNA was extracted using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). Animals were then genotyped using the Affymetrix Axiom OrcunSNP array (Affymetrix Inc., Santa Clara, CA, USA), which can analyse 199,692 SNPs. Low-quality SNPs were removed using the Axiom Analysis Suite and the PLINK v.1.9 software [22]. After filtering, 660 samples that had a call rate higher than 0.90 were retained and 139,922 SNPs remained for the subsequent analyses. Minor allele frequency within breeds and across breeds was not used to filter SNPs to avoid potential biases derived by the different number of animals analysed per breed.

Genetic diversity and population genomic parameters

Observed (H_o) and expected (H_e) heterozygosity and fixation index (F_{ST}) were calculated with the PLINK v.1.9 software [22]. F_{ST} values were based on the Hudson estimator, as this index is

independent of sample size [23]. The inbreeding coefficient of an individual (I) relative to the subpopulation (S) (F_{IS}) was calculated as $F_{IS} = 1-H_o/H_e$, [24]. The genetic distance between pairs of breeds was estimated as the average F_{ST} value across all SNPs [25, 26]. The resulting averaged F_{ST} values were used to build an F_{ST} matrix of size 15×15 that was then used to build a Neighbour-Joining (NJ) tree with the function "nj" in R v.4.0.4 based on 10,000 bootstrap replicates. Genetic differences between the 15 breeds were also evaluated using the genotyped SNPs with a multidimensional scaling (MDS) analysis, as implemented in PLINK v.1.9. Linkage disequilibrium (LD) was measured using r^2 for all SNP pairs on each chromosome using PLINK v.1.9 [22]. In addition, LD decay was estimated in bins of 50 kb to compare differences between and within breeds. Effective population size (N_e) at recent generations was computed using SNP data with the SNeP v.1.1 software based on default parameters [27]. Plots were generated in R v.4.0.4 [28].

Exploratory analysis of markers under selection

An exploratory analysis to detect outlier markers that could be under selection was performed using the *PCAdapt* package in R [29]. This analysis does not require prior knowledge on population structure and was performed on the merged dataset of the 15 rabbit breeds that consisted of 139,922 SNPs. Briefly, *PCAdapt* applies a Principal Components Analysis (PCA) aimed at selecting those components (K) explaining the greatest amount of variation, followed by a statistical test finalized at the detecting of outliers SNPs. As suggested by [30], the number of principal components K to work with was selected based on a Scree plot. For the detection of outliers SNPs, *PCAdapt* calculated a vector of z-scores $z_j = (z_{j1},..., z_{jK})$ obtained by regressing each j-th SNP by the k-th principal components via a multiple linear regression model [29]. Then, the Mahalanobis distance (D) statistical test was computed to detect ouliers SNPs as follows:

$$D_j^2 = (z_j - \overline{\mathbf{z}})^T \boldsymbol{\Sigma}^{-1} (z_j - \overline{\mathbf{z}}),$$

where $\boldsymbol{\Sigma}$ is the $(K \times K)$ covariance matrix of the *z*-scores and $\overline{z}$ is the vector of the K *z*-score means [31]. Mahalanobis distances were successively be transformed into *P*-values [29]. The threshold to identify outlier SNPs was defined based on a Bonferroni corrected *P*-value of 0.1.

Detection of signatures of selection

The F_{ST} analysis was further exploited to detect signatures of selection in the analysed rabbit breeds. F_{ST} single-marker-based analysis was performed by considering the markers at the extreme lower end of the distributions (99.95^{th} percentile of distribution). Since the OryCun2.0 rabbit genome assembly is not well defined (the N50 length for the contigs is 64,648 bp), many genomic regions are probably misplaced in the current assembly. Therefore, the identification of signatures of selection relied mainly on window-based analyses that could reduce the misassembly bias caused by small contigs.

F_{ST} was computed in 350-kb sliding genomic windows, with a step size of 100 kb. The choice of the best window size followed the method proposed by Rubin *et al.* [32]. Briefly, windows of variable sizes (from 50 to 500 kb) were evaluated for the number of windows with less than three SNPs. Window counts decreased asymptotically and stabilized after the 350-kb threshold (see Additional file 1 Table S1), resulting in 7951 genomic windows, with on average 17.5 ± 7 SNPs. Windows with less than three SNPs were not considered in the analyses. F_{ST} values were averaged across the SNPs in each genomic window. Two approaches were then applied to identify signatures of selection: (i) a single-breed approach to capture breed-specific features that could characterize each breed; and (ii) an approach based on groups of breeds to capture common features of several breeds.

For the first approach (single-breed approach), two methods were used to calculate window-based F_{ST} values:

(i) in Method 1 (M1), for each breed, the F_{ST} value of SNP was computed by comparing the given breed against all rabbits of the remaining $N-1$ breeds ($N = 15$), considered as a unique population [7] and averaged across all SNPs within a genomic window;

(ii) in Method 2 (M2), $N-1$ F_{ST} pairwise comparisons were performed for each breed by comparing the breed against one of the remaining $N-1$ breeds. The F_{ST} value of each genomic window for each comparison was computed as in M1 and then averaged across the $N-1$ F_{ST} pairwise comparisons for that window and breed [26, 33, 34].

In the second approach, groups of breeds were defined and contrasted in pre-defined pairwise comparisons. Groups of breeds were defined according to the following common features: (i) coat colours, (ii) coat colour patterns, (iii) body size, and (iv) commercial meat lines *vs* fancy breeds. Details of the groups of breeds and of the group-based F_{ST} pairwise analyses are summarised in Table S2 (see Additional File 1 Table S2).

Following Rubin *et al.* [19], we considered two stringency levels to identify the signatures of selection. Based on the most stringent level, signatures of selection were identified from genomic windows at the 99.8^{th} percentile of the distribution of the F_{ST} values. Suggestive signatures of selection were detected using a less stringent threshold that considered the 99.0^{th} percentile of the distribution.

To facilitate identification of candidate genes from the window-based approaches, considering potential assembly problems of the OryCun2.0 genome version (contig N50 = 64.648 kb), which could affect the precise genome position of annotated relevant genes, identified windows were further expanded by 200 kb at each side, i.e. about three times the contig N50 at each side [26, 35, 36]. Genomic regions were retrieved from the expanded windows for annotation and identification of overlaps between the two methods (M1 and M2) for each breed and combining results obtained

for all breeds. Data were graphically represented via Manhattan plots. Pipelines were developed either in Python or in R ("manhattan" function of the "qqman" library; [37]).

Genomic regions that displayed signatures of selection or suggestive signatures of selection were annotated with the Bedtools v.2.17 (https://bedtools.readthedocs.io/ [38]) by retrieving annotated protein coding genes from the OryCun2.0 NCBI's GFF file. The functional relevance of the genes annotated in the identified regions was evaluated based on a detailed analysis of the scientific literature, using the Gene Cards information [39] and the GWAS catalogue resource [40]. Over-representation analysis across sets of human traits was carried out for breed groups with Enrichr [41] via Fisher's exact test. The following libraries were interrogated: (i) the Gene Ontology (GO) Biological Process (PB) branch, (ii) the Kyoto Encyclopedia of Genes and Genomes (KEGG; Human), (iii) WikiPathways (Human), (iv) the MGI Mammalian Phenotype (Level 4) and (v) the GWAS catalogue. For each over-representation analysis, genes located in the 99.8^{th} percentile of genome windows were used as input set. Statistically enriched terms were defined if (i) they included at least two genes of the input set related to two or more of the top windows and (ii) had an adjusted *P*-value for enrichment less than 0.05.

Results

Within-breed genomic parameters

In total, 660 rabbits from 15 breeds that are characterized by different external features or purposes (coat colour and pattern, body size, fur type, and selected for meat production or for exhibitions/shows) were genotyped. Table 2 summarises some basic population genomic parameters for the analysed breeds. The average within-breed minor allele frequency (MAF) ± s.d. ranged from 0.178 ± 0.160 in the Ermine breed (a rare fancy breed) to 0.266 ± 0.152 in the Italian White breed (a selected meat line). Accordingly, the Ermine breed had the largest number of SNPs

(N = 47,409) with MAF lower than 0.05 and the Italian White breed had the largest number of informative SNPs (MAF ≥ 0.45) (N = 16,307). See also the MAF distributions in Fig. 1. Observed (H_o) and expected (H_e) heterozygosities ranged from 0.278 (Ermine) to 0.407 (Burgundy Fawn) and from 0.307 (Ermine) to 0.360 (Dwarf Lop), respectively. The value of the F_{IS} parameter was positive in four breeds (Coloured Dwarf, Dwarf Lop, Ermine and Rex), indicating non random mating in these breeds.

The LD decay between pairs of SNPs over a distance up to 300 kb is shown in Fig. 2a. LD dropped within the first 100 kb for all breeds and then reached approximate asymptotic but quite sizeable r^2 values. The Burgundy Fawn and Thuringian breeds had the highest level of LD and the smallest effective population size (N_e) across 50 generations [(Fig. 2b and Table 2) and (see Additional file 1 Table S3); N_e = 17 and N_e = 24, respectively]. It is worth mentioning that the results in these two breeds could be affected by their small sample sizes (less than 10). Among the other breeds, the smallest N_e was estimated for the Ermine, Italian Silver and Champagne d'Argent breeds (N_e = 32, 34 and 34, respectively) and the largest N_e for the Italian White (N_e = 134) and Checkered Giant (N_e = 124) breeds.

Genetic diversity and relationships among breeds

The level of population differentiation was estimated using the fixation index $F_{ST,}$ which was calculated as the average of pairwise breed comparisons per SNP using two methods (M1 and M2) (Table 2). The highest F_{ST} values were observed for the Burgundy Fawn and Ermine breeds for both methods (M1: 0.245 and 0.238; and M2: 0.284 and 0.282) and the lowest values for the Italian White breed (M1: 0.095; and M2: 0.216). Two groups of breeds could be identified based on the F_{ST} statistics: (i) a group of highly differentiated breeds, which included several fancy breeds, in addition to the Burgundy Fawn and Ermine breeds, i.e. the Belgian Hare, Champagne d'Argent,

Rhinelander, and Thuringian breeds; and (ii) a group of less differentiated breeds that included the three commercial breeds (Italian Silver, Italian Spot and Italian White) and several fancy breeds (Checkered Giant, Coloured Dwarf, Dwarf Lop, Giant Grey, Giant White and Rex).

In the single-breed pairwise F_{ST} analyses (single-SNP-based F_{ST} matrix), the highest F_{ST} values were found for the Burgundy Fawn *vs* Ermine breeds (F_{ST} = 0.351) and the lowest values for the Giant White *vs* Giant Grey breeds (F_{ST} = 0.119) (see Additional file 1 Table S4). The NJ tree based on F_{ST} distances (Fig. 2c) clustered the analysed breeds according to their morphological features and main purpose or type. The two dwarf breeds (Coloured Dwarf and Dwarf Lop) were clustered together with the Ermine breed, which is a small-sized breed. The three giant breeds (Checkered Giant, Giant Grey and Giant White) were also grouped together. The Italian Spotted and the Italian White breeds were grouped together in a cluster that, at a lower level, also included the Italian Silver and Champagne d'Argent breeds, which contributed to the genetic pool of the Italian Spotted breed. The Burgundy Fawn and Belgian Hare breeds were located in the same cluster as the Thuringian breed, which was positioned outside all other clusters. Similar results were obtained based on the F_{ST} window-based analysis (see Additional file 1 Table S5 and Additional file 2 Figure S1) that was also used to identify signatures of selection, as described below. The multidimensional scaling plots also clearly distinguished several breeds or groups of breeds, with a spatial disposition that resembled the NJ tree topography (Fig. 2d) and (see Additional file 2 Figure S2).

PCAdapt analysis: overview of markers under selection

According to Cattell's graphical rule, the last point before the curve flattens corresponds to the proper number of principal components that capture the population structure well [30]. Ten components (K = 10) were selected based on the Scree plot obtained from *PCAdapt* (see Additional file 2 Figure S3). This explorative analysis identified 280 outlier SNPs in the full dataset with all 15

rabbit breeds. The Manhattan plot obtained from the *PCAdapt* analysis (see Additional file 2 Figure S4) showed a few major peaks corresponding to several candidate genes that were also identified with the F_{ST} analyses (see below). These genes were located on *O. cuniculus* chromosome (OCU) 1 (*TYR*), OCU2 [*ligand dependent nuclear receptor corepressor like* (*LCORL*) / *non-SMC condensin I complex subunit G* (*NCAPG*)], OCU5 [*cadherin 13* (*CDH13*)], OCU8 [*endothelin receptor type B* (*EDNRB*)], OCU9 [*protein tyrosine phosphatase non-receptor type 2* (*PTPN2*)] and OCU14 (*LIPH*) (see Additional file 1 Table S6). The complete list of outlier markers, with overlapping or nearby annotated genes, is in Table S7 (see Additional file 1 Table S7). Figure 3 reports the positions of these outliers markers in the rabbit genome. However, this analysis based on all breeds did not enable the potential origins of the observed signals to be directly deduced (i.e. the breeds from which they derived or how they related to the phenotypic features of the genotyped animals), although the function of some of the genes (e.g. *TYR* determines several coat colours of the *albino* series and *LIPH* affects the coat structure of the Rex breed [15, 16, 18]) provided support that the correponding peaks contained signals that would be expected based on the phenotypic characteristics of some of the investigated breeds.

Signatures of selection detected using F_{ST} analyses of single breeds

In the analyses that compared one breed against all other breeds (Table 2) using Method 1, the window-averaged F_{ST} values across the genome ranged from 0.092 (Italian White *vs* all) to 0.231 (Burgundy Fawn *vs* all). Using Method 2, these same breeds also constituted the extremes (Italian White, F_{ST} = 0.206; Burgundy Fawn, F_{ST} = 0.287). F_{ST} values obtained from the two methods showed strong and significant correlation (r >0.7, P-value < 2E-16) (see Additional file 1 Table S8). Statistics on the single breed window-based F_{ST} analyses (Method 1 and Method 2) are reported in Table S9 (see Additional file 1 Table S9). The comparison of the F_{ST} values from the two methods suggested that Method 2 better highlighted genetic differences that may characterise

the tested breed against all other breeds, as its F_{ST} values were always higher than those obtained with Method 1.

In these analyses considering all 15 breeds, 210 and 1050 candidate selection signature windows were identified by each method based on the 99.8th (14 windows × 15 breeds) and 99.0th (70 windows × 15 breeds) percentile thresholds, respectively. The percentages of overlapping genomic windows obtained by Method 1 and Method 2 considering all breeds were 48% (99.8th percentile) and 54% (99.0th percentile), respectively. By merging the results obtained in the analysed breeds, Method 1 identified 190 (99.8th) and 869 (99.0th) unique candidate windows and Method 2 identified 193 (99.8th) and 883 (99.0th) unique candidate windows. After merging and expanding windows detected by the two methods by ±200 kb (Windows-M1 ∪ Windows-M2), 78 (99.8th) and 400 (99.0th) unique overlapping genomic regions were identified, comprising 47% (99.8th) and 51% (99.0th) of the overlapping genomic regions identifed by Method 1 and Method 2. Figure 3 shows the genome distribution of these regions (78 regions; 99.8th), combined for all breeds and the two methods, along with the distribution of 280 outlier SNPs obtained based on the *PCAdapt* analysis. The details for each breed are shown in Figure S5 (see Additional file 2 Figure S5) and the Manhattan plots obtained for each breed using Method 1 and Method 2 are in Figures S6 and S7 (see Additional file 2 Figures S6 and S7), respectively. The number of overlapping outlier regions identified using both methods ranged from two regions in the Italian Spotted and Giant Gray breeds to 11 regions in the Belgian Hare breed (see Additional File 1 Table S10). The complete lists of genome windows and regions identified from each of these analyses are in Tables S11 and S12 (see Additional file 1 Tables S11 and S12), respectively, and the list of annotated genes from the OryCun2.0 genome version is in Table S13 (see Additional file 1 Table S13). The list of genes obtained from the single-marker-based F_{ST} analysis with the single-breed approach (Method 1 and Method 2) are in Table S14 (see Additional file 1 Table S14) and the lists of markers obtained with

Methods 1 and Method 2 are in Tables S15 and S16 (see Additional file 1 Tables S15 and S16), respectively. A summary of the most relevant results obtained by the window-based single breed approach is in Table 3 and in Table S17 (see Additional file 1 Table S17). Several genomic regions that harbour strong candidate genes for their involvement in the determination of breed-specific features were identified with Method 1 and/or Method 2, indicating that the two methods could be considered, to some extent, as complementary in our approach based on outlier genomic windows detected with F_{ST} values against all other breeds averaged over SNPs in the window.

Some of these genes have already been reported to determine coat colour in some of the investigated breeds: *TYR* (identified in the Burgundy Fawn, Italian White, and Italian Spotted breeds), a gene located on OCU1 with variants that are responsible for the allele series of the *albino* locus and characterise the breeds for which this region was highlighted [15, 16]; *ASIP*, which is located on OCU4, was identified in Giant Grey rabbits, which carry a wild type allele at the *agouti* locus, and in Belgian Hare rabbits, which carry a missense mutation in this gene [13]; *MC1R*, a gene localized in the unplaced scaffold Un0267 [also known as GL018965 (ENSEMBL) or NW_003159591 (NCBI)], which was identified in the Rhinelander breed in which it determines the e^J allele of the *extension* locus that causes the classical tricolour phenotype of this breed [12].

Other coat colour genes that had not yet been associated with pigmentation phenotypes in rabbit, were found in several outlier genomic windows: the *EDNRB* gene, which was also identified with the *PCAdapt* analysis, is a gene involved in Hirschsprung disease type 2 in humans; it was located in an OCU8 outlier genomic window in the Rhinelander breed; the *melanocyte inducing transcription factor* (*MITF*) gene, which is involved in regulation of melanocyte development and in transcription of melanogenesis enzyme genes, was located in an OCU9 F_{ST} extreme window in the Giant White breed; the *endothelin receptor type A* (*EDNRA*) gene, which is suggested to be

involved in loss of pigmentation in goats [42], was located in an extreme F_{ST} region of OCU15 in the Italian White and Thuringian breeds; and the *OCA2 melanosomal transmembrane protein* (*OCA2*) gene, which is the homolog of the mouse *p* (pink-eyed dilution) gene that is involved in type 2 oculocutaneous albinism, was located in a OCU17 outlier region in the Checkered Giant breed.

Several additional signatures of selection were identified in genomic regions that contained genes already known to affect other external traits in rabbits. For example, in Rex rabbits, a signature of selection was identified in a region of OCU14 that harbours the *lipase H* (*LIPH*) gene, which is responsible for the effect of the Rex^1 locus on coat structure [18]. A genomic window on OCU2 that harbours the *ligand dependent nuclear receptor corepressor like* (*LCORL*) and the *non-SMC condensin I complex subunit G* (*NCAPG*) genes, which are known to affect body size in several mammals (e.g. [43-48]) and was previously identified in a signature of selection in the rabbit genome [19], was detected in the Dwarf Lop and Ermine breeds. The genomic regions with the *LIPH* and *LCORL/NCAPG* genes were also identified with the *PCAdapt* analysis. Another region on OCU13 that was previously detected by Carneiro *et al.* [19] and harbours the *collagen type XI alpha 1 chain* (*COL11A1*) gene was identified in the Coloured Dwarf and Ermine breeds. This gene is essential in skeletal morphogenesis and variants have been associated with body height in humans [49, 50].

Other candidate genes that potentially affect body size were located in several other extreme genomic windows for some of the breeds investigated. Among these novel candidate genes (reported for the first time in this study), *collagen type II alpha 1 chain* (*COL2A1*), located in an extreme genomic region of an unassembled scaffold (Un0251) in the Dwarf Lop breed, has been reported to cause a wide spectrum of skeletal disorders in mammals, including achondrogenesis

type II in humans [51-53] and the bulldog-type dwarfism in cattle [54-58]. Another region on OCU18 that carries a candidate gene involved in body size (*G protein-coupled receptor kinase 5*; *GRK5*) was also identified in the Dwarf Lop breed. Variability in the *GRK5* gene is strongly associated with body height in humans [50].

The complexity of the genetic factors that affect dwarfism and small body size was also evidenced by some differences among the three dwarf/small body size breeds. For example, a genomic window on OCU13 (position 38.50-38.85 Mb) that contains a body size-related gene (*GATA zinc finger domain containing 2B*, or *GATAD2B*; [59]) was identified in the Coloured dwarf breed. Close to this window, another window (position 37.10-37.45 Mb) was identified in this same breed (99.8th percentile) and in the Ermine breed (99.0th percentile). This window includes several genes associated with body size and height in humans: *Rho/Rac guanine nucleotide exchange factor 2* (*ARHGEF2*) associated with body size in children; *mex-3 RNA binding family member A* (*MEX3A*) associated with adult body size; *zinc finger and BTB domain containing 7B* (*ZBTB7B)*, associated with adult body size and birth weight; and *DC-STAMP domain containing 1* (*DCST1*) associated with height [50, 60, 61].

F_{ST} analyses between groups of breeds

To identify additional signatures of selection and to confirm regions identified in the single-breed analyses, some breeds were grouped together according to common features (shared coat colours, colour patterns, body size and use/specialization). Detailed statistics on the window-based F_{ST} analyses obtained for different groups of breeds are in Table S18 (see Additional file 1 Table S18), and the complete list of the genomic windows and regions identified with the six pairwise group comparisons is in Table S19 [see Additional file 1 Table S19). The most relevant genes identified in these analyses are in Table 4. The results obtained from the F_{ST} single-marker-based analysis for the

groups of breeds are in Tables S20 and S21 (see Additonal file 1 Tables S20 and S21). The annotated Miami plots (Fig. 4) include the most relevant results obtained from the window-based and single-marker-based F_{ST} analyses.

By grouping the two albino breeds (Giant White and Italian White), we identified a peak on OCU1 that includes the *TYR* gene (among the top 99.8th outlier genomic windows), as expected according to previous studies [15, 16]. In this comparison, an outlier genomic windows on OCU15 harbours the *EDNRA* gene [42], confirming our novel result reported in the single-breed F_{ST} analysis. The analysis that was based on the two silver/greying breeds (Champagne d'Argent and Italian Silver), which was mainly aimed at identifying genome regions that might be involved in their peculiar progressive graying of the hairs, identified outlier windows on OCU9 (two windows, 43.05-43.40 Mb and 46.20-46.55 Mb), on OCU8 (two windows from 64.4 to 65.10 Mb), on OCU13 (two windows from 115.15 to 115.85 Mb and two from 123.20 to 125.65 Mb), on OCU1 (a window from 78.05 to 78.40 Mb), on OCU12 (a window from 74.90 to 75.25 Mb), and on unassembled scaffolds (Un0088, Un0265, Un0329 and Un0513). These outlier genomic regions contain several genes that could be involved in defining this coat colour related trait (Table 4) since they play roles in the regulation of cell death (*ADNP2*; scaffold Un513) and in the control of aging in hair follicle stem cells (*NFATC1*; scaffold Un0329) [62-64]. Analysis of the two checkered spotted breeds (Checkered Giant and Rhinelander) as a group identified a signature of selection on OCU4 among the top outlier regions, which was located close to the *KIT ligand* (*KITLG*) gene. KITLG is the ligand of the tyrosine-kinase receptor encoded by the *KIT* gene that is involved in cell migration processes. As the classical spotted phenotype in these breeds has been associated with heterozygotes for a variant in the *KIT* gene [17], *KITLG* could contribute to the specific positioning of the pigmented skin areas of the checkered pattern. Notably, no signature of selection was evident in the *KIT* gene region in all comparisons, likely because the *English spotted* phenotype derives

from heterozygosity at this locus [17], which might have reduced F_{ST} values for this region. The genotype information and the observed heterozygosity for variants in the *KIT* gene region are in Tables S22 and S23 (see Additional file 1 Tables S22 and S23), respectively. The mean observed heterozygosity for the two spotted breeds was (0.567 for Rhinelander and 0.626 for Checkered Giant) was close to that expected according to our previous study on this gene [17].

When the three dwarf/small breeds were considered together, a signature of selection in the OCU2 region that harbors the *LCORL* and *NCAPG* genes was again observed, in addition to regions on several other chromosomes (OCU1, 12, 14 and 18) and unassigned scaffolds (Un0030, Un0044 and Un0076), most of which include genes that were previously associated with body size and height in humans or other species. For example, the region on OCU18 includes the *glutamate ionotropic receptor delta type subunit 1* (*GRID1*) gene and the four outlier continuous windows on the unassigned scaffold Un0030 include the *neurotrophic receptor tyrosine kinase 2* (*NTRK2*) and the *FERM domain containing 3* (*FRMD3*) genes, which have been associated with body weight in humans [50]. Other regions were identified by Carneiro *et al.* [19] to be associated with small body size and dwarfism in rabbits. Among these regions, in our analysis, in addition to the *LCORL*/*NCAPG* region, we also confirmed the region on OCU7 containing the *HOXD* gene cluster and the region on OCU13 containing the *COL11A1* gene but only at the less stringent threshold of 99.0^{th} percentile.

The top 99.8^{th} genome windows that were identified when grouping the three giant breeds (Checkered Giant, Giant Grey and Giant White) were located on nine chromosomes (OCU1, 2, 3, 5, 7, 12, 13, 14 and 16) and on several unassigned scaffolds (Un0030 and Un0366). These windows include several genes related to growth traits and regulators of developmental processes. For example, the most extreme window, on OCU12, harbours the *bone morphogenetic protein 5* (*BMP5*) gene, which plays a key role in skeletal morphogenesis [65], and the *collagen type XXI*

alpha 1 chain (*COL21A1*) gene, which has been associated with body size in humans [61]. The next top three windows, on OCU5, 16 and 2, harbour other genes that were previously associated with body size and height [61] [*cadherin 13* (*CDH13*), *solute carrier family 30 member 10* (*SLC30A10*) and *methionine sulfoxide reductase A* (*MSRA*), respectively). The signal on OCU5, harbouring the *CDH13* gene, was also detected in the *PCAdapt* analysis (see Additional file 2, Figure S4). It is also interesting to note that another genome window on OCU4 that was identified at the 99.0th percentile threshold, contained the *HMGA2* gene, which has been consistently reported to be associated with stature and body size in humans and other species [47, 66-68]. A large deletion in this gene is also responsible for a form of dwarfism in rabbit [19]. The genotype information and the observed heterozygosity values for variants in the *HMGA2* gene region are in Tables S24 and S25 [see Additional file 1 Tables S24 and S25), respectively. The mean observed heterozygosity was higher in the dwarf breeds (0.380 in Dwarf Lop, 0.731 in Coloured Dwarf and 0.593 in Ermine) than in the Giant breeds (0.127 in Giant White, 0.123 in Giant Grey and 0.124 in Checkered Giant).

The signatures of selection that were identified by combining the three meat rabbit breeds overlapped partially with those obtained in the comparisons based on groups of rabbits of similar body size. Outlier regions were identified on nine chromosomes and on two unplaced scaffolds (see Additional file 1 Table S19), which included a number of novel candidate genes. The top 99.8th genome window on OCU2 includes the *mitochondrial ribosomal protein L33* (*MRPL33*) gene, which has been reported to be a candidate gene for muscle development in ducks [69]. The second top genomic window on OCU15 contains the *coiled-coil serine rich protein 1* (*CCSER1*), which has been suggested to be involved in fat deposition in pigs [70]. Another important candidate gene in a window that was identified on OCU9, *protein tyrosine phosphatase non-receptor type 2* (*PTPN2*), confirms the results of the *PCAdapt* analysis (see Additional file 2, Figure S4). This gene is a member of the PTP family which includes signaling molecules that regulate cell growth,

differentiation, the mitotic cycle, and oncogenic transformation [71]. Another gene that controls cell differentiation (*neurotrophic receptor tyrosine kinase 2*; *NTRK2*) is located in an identified genome window on the unassigned scaffold Un0030. In humans, mutations in this gene are associated with severe obesity [72].

The over-representation analysis based on genes that were located in the extreme outlier windows (99.8th percentile) and were detected for each of the five tested groups of breeds identified three over-represented biological features (one biological process and two human phenotypes) for comparisons involving the two silver/greying breeds and the two checkered spotted breeds (see Additional file 1 Table S26). The enriched terms did not point to any relevant characteristics related to the targeted phenotypes.

Discussion

Domestication and the subsequent directional artificial and natural selection, along with several other genetic events, have shaped the genome of all domestic animals and differentiated many breeds and populations within species. These genetic resources, which were generally constituted relatively recently, have been used to dissect the genetic mechanisms that determine extreme morphological and physiological features in several species (e.g. [73-77]). The large number of rabbit breeds represents a unique and mostly unexplored resource, which can help to dissect the genetic architecture of exterior traits in this lagomorph species and, as a mirror, also in other mammals [1]. The rabbit is used as an animal model for applied and basic biological studies, mainly where the rodents have demonstrated several limits [78].

In this study, we investigated different fancy rabbit breeds that have characteristic phenotypes, i.e. coat colours, colour patterns, and body size and morphology. In addition, three pure breeds used to produce a three-way crossbred commercial meat line, were analysed.

As a proof of concept to demonstrate that the applied methodologies can capture recent signatures of selection, several of the results that we obtained highlight previously identified genes that affect coat colour (*ASIP*, *MC1R* and *TYR*; [11-16]), coat structure (*LIPH*; [18]), and body size (*LCORL/NCAPG*, *COL11A1* and *HOXD*; [19]) in rabbit. Other identified regions of signatures of selection also contain many novel candidate genes for phenotypic variation among the analysed breeds. Thus, complementing the candidate gene approach of previous studies in rabbit that were successful mainly for monogenic traits, the hitchhiking comparative mapping approach across breeds that we applied here, can provide a more complete view of the genetic elements that determine the main phenotypic differences in this species. In particular, many signatures of selection pointed to novel candidate genes for body size and could contribute to understand the genetic mechanisms that affect similar traits and phenotypes in other mammals. In addition, several other genes that affect coat colour and were not previously known to be involved in this phenotype in rabbits were identified in regions containing signatures of selection. Overall, 309 unique genomic regions, obtained by combining the results from single-breed analyses and analyses of groups of breeds, were identified as implicated in the most relevant genetic differences between the breeds investigated.

In spite of the insights obtained from our results, it is important to note a few limitations and how to overcome them, at least in part. Based on how the rabbit breeds have been constituted, with bottleneck effects, genetic drift, and introgression potentially playing important roles, a genome region that possesses a pattern of differentiation between breeds does not necessarily prove that the region has been under selection and has functional roles. The function of the annotated genes in the identified regions, which usually derives from what is known from other species, can be useful in some cases to interpret the obtained results. However, not all genes are functionally well described in the literature, which leaves many uncertainties, especially for complex traits or when multiple

pathways are involved in the breeds' differential phenotypes. Therefore, to extract meaningful biological features, for group-based results, we applied enrichment approaches that attempted to improve the insights into the complexity of body size and selection for productive performances. However, this analysis was not very informative, probably due to the heterogeneity of the biological mechanisms, with involvement of a large number of genes that are present in the identified genomic regions. The inclusion of additional animals per breed, additional breeds, and other breed populations, e.g. sampled in more countries (rabbit breeds are, in many cases, named according to common features but the genetic history of each national stock can differ [8, 9]) may be able to improve the comparative resolution that would be needed to reduce the noise generated by stochastic effects and demographic and specific population-derived perturbations. Other statistical methodologies that have been proposed to identify signatures of selection, along with additional genomic information (i.e. whole-genome sequencing data), could be useful to complement the approaches and data used in this study and to capture other selection signals [6, 19]. In addition, several selection signals were detected in unassigned scaffolds. Thus, improvement of the assembly of the rabbit reference genome is needed to better detect, at the chromosome level, signatures of selection that could have been fragmented or missed in the current study.

Our study further improves the complex puzzle of coat colour genetics in the rabbit, which was already drafted in its main framework by previous studies on candidate genes, as mentioned [1]. Signatures of selection were detected in genomic windows containing *EDNRB*, *EDNRA*, *MITF*, and *OCA2* that, to date, were not reported to affect coat colour in rabbits, although the role of these genes on coat colour is well known in several other species. We previously excluded a polymorphism in the *EDNRB* gene to be associated with the *English spotting* locus [79] but, according to our results from this study, this gene may have a role in modifying or regulating the position/extension of the bicolor (red and black) spotted patterns over the unpigmented background

of the Rhinelander breed. Similar secondary roles in defining the main coat colours in the Giant White, Italian White, Thuringian, and Checkered Giant breeds, or the result of other hitchhiking effects may explain the identification of highly differentiated regions in these breeds that include *MITF*, *EDNRA,* and *OCA2*. However, the role of some of these genes in the coat colour of these breeds should be further analysed. The silvering or hereditary greying of hair, which is expected to be caused by a recessive *si* allele at the *Silver* locus (reviewed in [1]), was investigated by comparing the Champagne d'Argent and Italian Silver (a breed derived from Champagne d'Argent) breeds. No obvious candidates were reported in the outlier genomic windows that were identified in this analysis, although the function of some of the genes in these regions was directly or indirectly consistent with this coat colour phenotype. More detailed analyses, including whole-genome sequencing data, could be useful to fine map the candidate region that harbors the causative mutation(s) for this coat colour locus.

The complexity and heterogeneity of the signatures of selection that were detected in the comparisons involving the dwarf/small body size breeds, the giant breeds, and the three meat breeds indicate that a large number of genes with small effects are involved in determining body size and production performances in the rabbit, similarly to what is reported in other mammals (e.g. [42, 44-47, 49]. A few major genes (*LCORL*/*NCAPG*, the *HOXD* gene family, and *COL11A1*), which have alleles that might contribute to reducing body size in rabbit, were identified in this study, confirming the results already reported in a previous study based on Netherland Dwarf rabbits [19]. No signal in the *HMGA2* gene region was identified in the dwarf/small body size breeds that we analysed in our study, although a large deletion that overlaps the promoter region and the first three exons of this gene has been reported to be the causative mutation of a form of dwarfism [19]. This could be due to viable dwarf rabbits being heterozygous for the mutated allele, since the homozygous state is usually lethal [19]. Such a heterozygous condition that does not create extreme

allele differences, may not have been captured by the F_{ST} analysis used in our study. However, this interpretation is supported by the results obtained from the observed heterozygosity analysis for variants within the *HMGA2* gene region, since the average level of heterozygosity was close to 0.5. It is also possible that, in our dwarf breeds, other loci cause this extreme reduction in size. For example, in the Dwarf Lop breed, we identified a strong signature of selection in the region of the *COL2A1* gene, which is known to be involved in a broad spectrum of skeletal defects, including dwarfism [51-54]. Identification of selection signature in the *HMGA2* gene region in the giant breeds was interesting and was probably due to the presence of alternative alleles for this gene that increase body size. In general, our results based on the comparisons between breeds of extreme body size demonstrate the complexity of the genetic factors that affect this phenotype, which is usually referred as a breed-specific trait in rabbit. Body size, body structure, and stature are quantitative traits for which some major genes have been identified, as already demonstrated in several other mammals [47, 48, 80-82].

Several other interesting candidate genes were detected by comparing the genomes of meat breeds with fancy breeds, which are not selected for muscle growth and performance traits. The most interesting region, which was also found with the *PCAdapt* analysis, contains the *PTPN2* gene, which is involved in several regulatory and cell differentiation mechanisms [71]. Other studies are needed to characterise the function of this gene and its role in growth and performance traits in rabbit and other livestock species.

Conclusions

This study is the first genome-wide analysis of signatures of selection using high-density SNP genotype data in rabbits. Fifteen rabbit breeds with divergent coat colours and colour patterns, body sizes and uses/specializations were analysed. We detected several regions with significant

signatures of selection, which open new avenues for further investigations to characterize the identified candidate genes and confirm their causative role. In particular, this study contributed to enlarge the number of candidate genes for body size and coat colour in this multi-purpose species. Results showed that body size is also a complex trait in rabbit, and maybe determined by the effects of many genes, among which some could have a potential major effect based on the analysis of signatures of selection. Other investigations with additional breeds and populations and by using different statistical approaches are needed to expand this analysis of signatures of selection in the rabbit genome. Our results will be useful to better understand rabbit domestication, which appears to be nested with the constitution of the main breeds, at least for the most recent developments. These breeds represent important genetic resources for which complete characterization at the genome level has only just started.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author upon reasonable request and can be shared after an agreement on their use with University of Bologna and ANCI.

Competing interests

The authors declare that they have no competing interests.

Funding

The study was funded by the PSRN (Programma di Sviluppo Rurale Nazionale) Cun-Fu and Cun-Fu 2 projects (co-funded by the European Agricultural Fund for Rural Development of the European Union and by the Italian Ministry of Agriculture, Food and Forestry - MiPAAF) and by the University of Bologna RFO 2019 programme.

Authors' contributions

LF conceived and designed the study and obtained funding. MB, SB and GS conducted bioinformatics analyses. MS and RN provided samples and resources. All authors contributed to data interpretation. LF and SB wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

The authors thank rabbit breeders for their collaboration in this study.

References

1. Fontanesi L. Rabbit genetic resources can provide several animal models to explain at the genetic level the diversity of morphological and physiological relevant traits. Appl Sci (Basel). 2021;11:373.

2. Zeder MA. The domestication of animals. J Anthropol Res. 2012;68:161–90.

3. Zeuner FE. A history of domesticated animals. Evanston: Harper & Row Publishers; 1963.

4. Callou C. De la garenne au clapier. Étude archéozoologique du lapin en Europe occidentale. Mémoires du Muséum National d'Histoire Naturelle. 2002.

5. Fontanesi L, Utzeri VJ, Ribani A. The evolution, domestication and world distribution of the European rabbit (*Oryctolagus cuniculus*). In: Fontanesi L, editor. The genetics and genomics of the rabbit. Wallingford: CAB International; 2021. p. 1–22.

6. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Martinez Barrio A, et al. Rabbit

genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 2014;345:1074–79.

7. Alves JM, Carneiro M, Afonso S, Lopes S, Garreau H, Boucher S, et al. Levels and patterns of genetic diversity and population structure in domestic rabbits. PLoS One. 2015;10:e0144687.

8. Whitman BD. Domestic rabbits and their histories: Breeds of the world. Overland Parks: Leathers Publishing; 2004.

9. Boucher S, Garreau H, Boettcher P, Bolet G. Rabbit breeds and lines and genetic resources. In: Fontanesi L, editor. The genetics and genomics of the rabbit. Wallingford: CAB International; 2021. p. 23–37.

10. Searle, A. G. Comparative genetics of coat colour in mammals. London: Logos Press Limited; 1968.

11. Fontanesi L, Tazzoli M, Beretti F, Russo V. Mutations in the melanocortin 1 receptor (*MC1R*) gene are associated with coat colours in the domestic rabbit (*Oryctolagus cuniculus*). Anim Genet. 2006;37:489–93.

12. Fontanesi L, Scotti E, Colombo M, Beretti F, Forestier L, Dall'Olio S, et al. A composite six bp in-frame deletion in the melanocortin 1 receptor (*MC1R*) gene is associated with the Japanese brindling coat colour in rabbits (*Oryctolagus cuniculus*). BMC Genet. 2010;11:59.

13. Fontanesi L, Forestier L, Allain D, Scotti E, Beretti F, Deretz-Picoulet S, et al. Characterization of the rabbit agouti signaling protein (*ASIP*) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. Genomics. 2010;95:166–75.

14. Letko A, Ammann B, Jagannathan V, Henkel J, Leuthard F, Schelling C, et al. A deletion spanning the promoter and first exon of the hair cycle-specific *ASIP* transcript isoform in black and tan rabbits. Anim Genet. 2020;51:137–40.

15. Aigner B, Besenfelder U, Müller M, Brem G. Tyrosinase gene variants in different rabbit strains. Mamm Genome. 2000;11:700–2.

16. Utzeri VJ, Ribani A, Schiavo G, Fontanesi L. Describing variability in the tyrosinase (*TYR*) gene, the albino coat colour locus, in domestic and wild European rabbits. Ital J Anim Sci. 2021;20:181–7.

17. Fontanesi L, Vargiolu M, Scotti E, Latorre R, Faussone Pellegrini MS, Mazzoni M, et al. The *KIT* gene is associated with the English spotting coat color locus and congenital megacolon in Checkered Giant rabbits (*Oryctolagus cuniculus*). PLoS One. 2014;9:e93750.

18. Diribarne M, Mata X, Chantry-Darmon C, Vaiman A, Auvinet G, Bouet S, et al. A deletion in exon 9 of the *LIPH* gene is responsible for the rex hair coat phenotype in rabbits (*Oryctolagus cuniculus*). PLoS One. 2011;6:e19281.

19. Carneiro M, Hu D, Archer J, Feng C, Afonso S, Chen C, et al. Dwarfism and altered craniofacial development in rabbits is caused by a 12.1 kb deletion at the *HMGA2* locus. Genetics. 2017;205:955–65.

20. Fontanesi L. Genetics and molecular genetics of morphological and physiological traits and inherited disorders in the European rabbit. In: Fontanesi L, editor. The Genetics and Genomics of the Rabbit. Wallingford: CAB International; 2021. p. 120–162.

21. Bertolini F, Schiavo G, Scotti E, Ribani A, Martelli PL, Casadio R, et al. High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. Anim Genet. 2014;45:304–7.

22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.

23. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. Genome Res. 2013;23:1514–1521.

24. Nei M, Chesser RK. Estimation of fixation indices and gene diversities. Ann Hum Genet. 1983;47:253–9.

25. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NHC, Zody MC, Anderson N, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet. 2007;39:1321–8.

26. Bovo S, Ribani A, Muñoz M, Alves E, Araujo JP, Bozzi R, et al. Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. Genet Sel Evol. 2020;52:33.

27. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 2015;6:109.

28. R core team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

29. Luu K, Bazin E, Blum MG. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Res. 2017;17:67–77.

30. Cattell RB. The scree test for the number of factors. Multivariate Behav Res. 1966;1:245–76.

31. Maronna RA, Zamar RH. Robust estimates of location and dispersion for high-dimensional datasets. Technometrics. 2002;44:307–17.

32. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464:587–91.

33. Guo J, Tao H, Li P, Li L, Zhong T, Wang L, et al. Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. Sci Rep. 2018;8:10405.

34. Muñoz M, Bozzi R, García-Casco J, Núñez Y, Ribani A, Franci O, et al. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. Sci Rep. 2019;9:13546.

35. Cheruiyot EK, Bett RC, Amimo JO, Zhang Y, Mrode R, Mujibi FD. Signatures of selection in admixed dairy cattle in Tanzania. Front Genet. 2018;9:607.

36. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. Signatures of selection and environmental adaptation across the goat genome post-domestication. Genet Sel Evol. 2018;50:57.

37. Turner S. qqman: an R package for visualizing GWAS results using Q–Q and Manhattan plots. J Open Source Softw. 2018;3:731.

38. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

39. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: From gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics. 2016;54:1.30.1-1.30.33.

40. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics. Nucleic Acids Res. 2019;47:D1005–12.

41. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.

42. Menzi F, Keller I, Reber I, Beck J, Brenig B, Schütz E, et al. Genomic amplification of the caprine EDNRA locus might lead to a dose dependent loss of pigmentation. Sci Rep. 2016;6:28438.

43. Pryce JE, Hayes BJ, Bolormaa S, Goddard ME. Polymorphic regions affecting human height also control stature in cattle. Genetics. 2011;187:981–4.

44. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci USA. 2012;109:19529–36.

45. Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, et al. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PLoS One. 2012;7:e37282.

46. Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. Genet Sel Evol. 2015;47:66.

47. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50:362–7.

48. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat Commun. 2019;10:1489.

49. Li Y, Lacerda DA, Warman MA, Beier DR, Yoshioka H, Ninomiya Y.et al. A fibrillar collagen gene, *Col11a1*, is essential for skeletal morphogenesis. Cell. 1995;80:423–30.

50. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;104:65–75.

51. Vissing H., D'Alessio M, Lee B, Ramirez F, Godfrey M, Hollister DW. Glycine to serine substitution in the triple helical domain of pro-α1 (II) collagen results in a lethal perinatal form of short-limbed dwarfism. J Biol Chem. 1989;264:18265–7.

52. Körkkö J, Cohn DH, Ala-Kokko L, Krakow D, Prockop DJ. Widely distributed mutations in

the *COL2A1* gene produce achondrogenesis type II/hypochondrogenesis. Am J Med Genet. 2000;92:95–100.

53. Forzano F, Lituania M, Viassolo V, Superti-Furga A, Wildhardt G, Zabel B, et al. A familial case of achondrogenesis type II caused by a dominant *COL2A1* mutation and "patchy" expression in the mosaic father. Am J Med Genet. 2007;143A:2815–20.

54. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014;46:858–65.

55. Reinartz S, Mohwinkel H, Sürie C, Hellige M, Feige K, Eikelberg D, et al. Germline mutation within *COL2A1* associated with lethal chondrodysplasia in a polled Holstein family. BMC Genomics. 2017;18:762.

56. Häfliger IM, Behn H, Freick M, Jagannathan V, Drögemüller C. A *COL2A1* de novo variant in a Holstein bulldog calf. Anim Genet. 2019;50:113–4.

57. Jacinto JGP, Häfliger IM, Letko A, Drögemüller C, Agerholm JS. A large deletion in the *COL2A*1 gene expands the spectrum of pathogenic variants causing bulldog calf syndrome in cattle. Acta Vet Scand. 2020;62:49.

58. Jacinto JG, Häfliger IM, Gentile A, Drögemüller C, Bolcato M. A 6. 7 kb deletion in the *COL2A1* gene in a Holstein calf with achondrogenesis type II and perosomus elumbis. Anim Genet. 2021;52:244–5.

59. Akiyama M, Ishigaki K, Sakaue S, Momozawa Y, Horikoshi M, Hirata M, et al. Characterizing rare and low-frequency height-associated variants in the Japanese population. Nat Commun. 2019;10:4393.

60. Rüeger S, McDaid A, Kutalik Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. PLoS Genet. 2018;14:e1007371.

61. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. BMJ. 2020;369:m1203.

62. Kushnir M, Dresner E, Mandel S, Gozes I. Silencing of the ADNP-family member, *ADNP2*, results in changes in cellular viability under oxidative stress. J Neurochem. 2008;105:537–45.

63. Keyes BE, Segal JP, Heller E, Lien W-H, Chang C-Y, Guo X, et al. *Nfatc1* orchestrates aging in hair follicle stem cells. Proc Natl Acad Sci USA. 2013;110:E4950–9.

64. Fridmanis D, Roga A, Klovins J. ACTH receptor (MC2R) specificity: What do we know about underlying molecular mechanisms? Front Endocrinol (Lausanne). 2017;8:13.

65. Kingsley DM, Bland AE, Grubber JM, Marker PC, Russell LB, Copeland NG, et al. The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGFβ superfamily. Cell. 1992;71:399–410.

66. Weedon MN, The Diabetes Genetics Initiative, Lettre G, Freathy RM, Lindgren CM, Voight BF, et al. A common variant of *HMGA2* is associated with adult and childhood height in the general population. Nat Genet. 2007;39:1245–50.

67. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. PLoS Biol. 2010;8:e1000451.

68. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, et al. Four loci explain 83% of size variation in the horse. PLoS One. 2012;7:e39929.

69. Xu T, Huang W, Zhang X, Ye B, Zhou H, Hou S. Identification and characterization of genes related to the development of breast muscles in Pekin duck. Mol Biol Rep. 2012;39:7647–55.

70. Xue Y, Li C, Duan D, Wang M, Han X, Wang K, et al. Genome-wide association studies for growth-related traits in a crossbreed pig population. Anim Genet. 2021;52:217–22.

71. Huang Z, Liu M, Li D, Tan Y, Zhang R, Xia Z, et al. PTPN2 regulates the activation of KRAS and plays a critical role in proliferation and survival of KRAS-driven cancer cells. J Biol Chem. 2020;295:18343–54.

72. Yeo GS, Connie Hung CC, Rochford J, Keogh J, Gray J, Sivaramakrishnan S, et al. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci. 2004;7:1187-9.

73. Bovine HapMap Consortium. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009;324:528–32.

74. Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, et al. Tracking footprints of artificial selection in the dog genome. Proc Natl Acad Sci USA. 2010;107:1160–5.

75. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, et al. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet. 2013;9:e1003211.

76. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. Signatures of selection and environmental adaptation across the goat genome post-domestication. Genet Sel Evol. 2018;50:57.

77. Qanbari S, Rubin C-J, Maqbool K, Weigend S, Weigend A, Geibel J, et al. Genetics of adaptation in modern chicken. PLoS Genet. 2019;15:e1007989.

78. Miller I, Rogel-Gaillard C, Spina D, Fontanesi L, de Almeida AM. The rabbit as an experimental and production animal: from genomics to proteomics. Curr Protein Pept Sci. 2014;15:134–45.

79. Fontanesi L, Vargiolu M, Scotti E, Mazzoni M, Clavenzani P, De Giorgio R, et al. Endothelin receptor B (*EDNRB*) is not the causative gene of the English spotting locus in the domestic rabbit (*Oryctolagus cuniculus*). Anim Genet. 2010;41:669–70.

80. Kemper KE, Visscher PM, Goddard ME. Genetic architecture of body size in mammals. Genome Biol. 2012;13:244.

81. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46:1173–86.

82. Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, et al. Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Res. 2013;23:1985–95.

Figures

Figure 1 Distribution of minor allele frequencies (MAF).

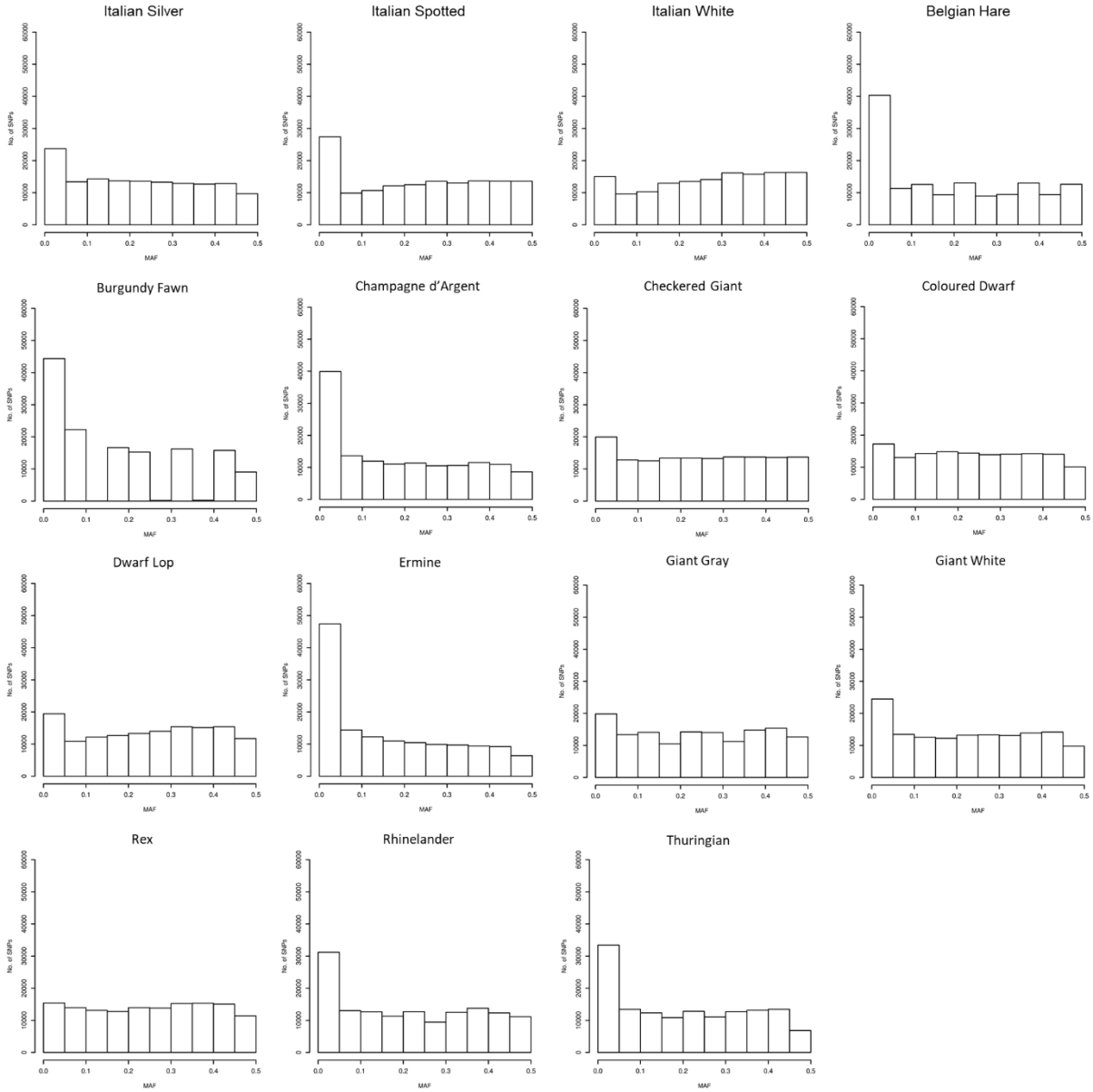


Figure 2 Results of population genomic analyses of 15 rabbit breeds.

(**a**) Linkage disequilibrium (LD) decay; (**b**) Effective population size (N_e); (**c**) Neighbor Joining (NJ) tree (next to the branches, the bootstrap test values are indicated in red, expressed as percentage over 10,000 replicates) and (**d**) 3D multidimensional scaling (MDS) plot.

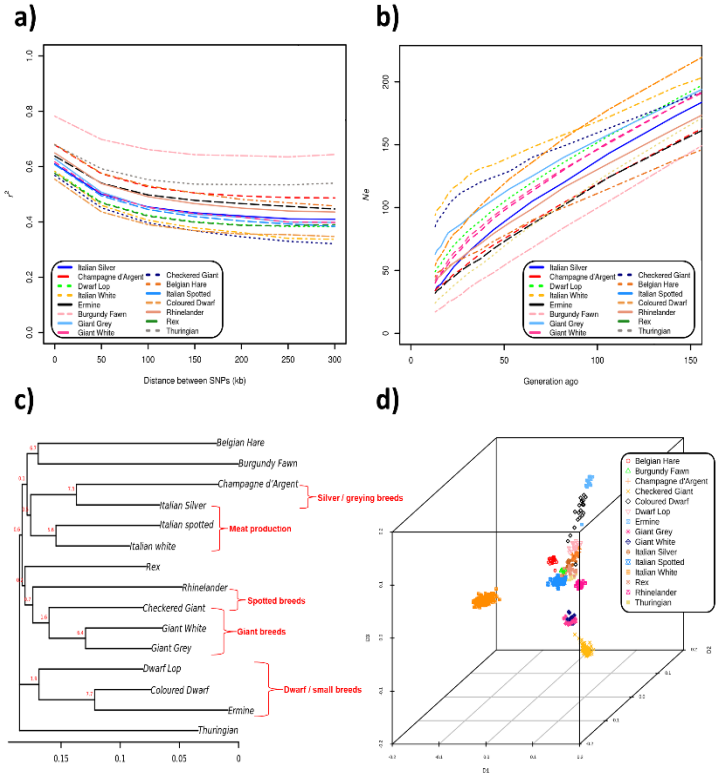


Figure 3 Genomic regions with signatures of selection that were identified using window-based F_{ST} analyses in the single-breed approach (Method 1 and Method 2) and *PCAdapt* analysis.

Outlier windows (99.8th percentile threshold) were expanded and merged into genomic regions. Only the assembled autosomes are presented and unassembled scaffolds are not included.

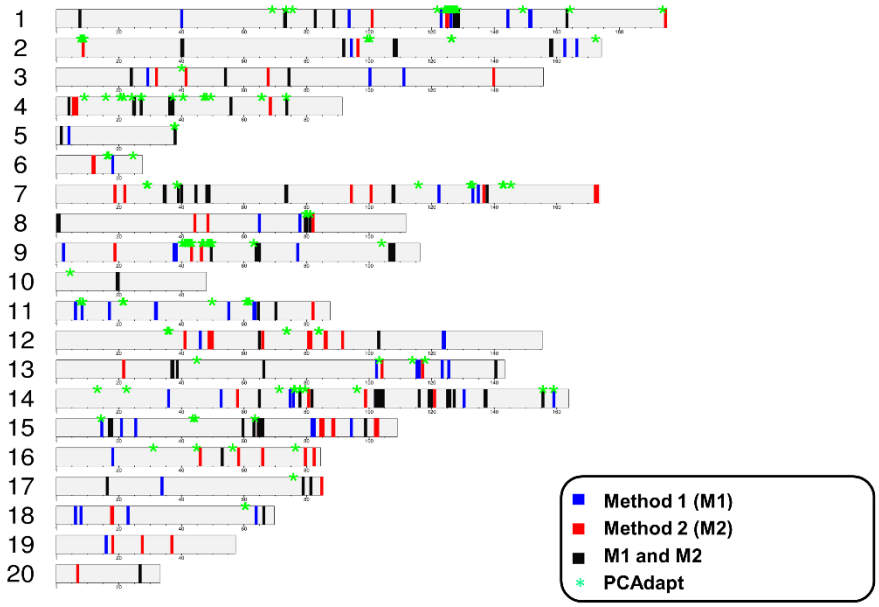
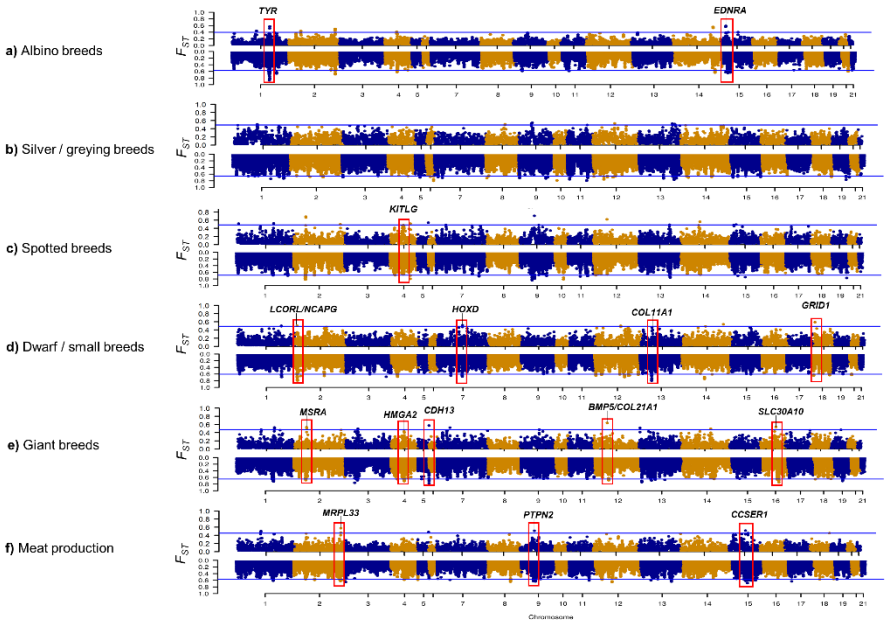


Figure 4 Miami plots of the genome-wide window-based (top) and single-marker-based (bottom) FST analyses of groups of breeds.

(**a**) Albino breeds; (**b**) Silver/greying breeds (genes are not reported as they map to scaffolds); (**c**) Spotted breeds; (**d**) Dwarf/small breeds; (**e**) Giant breeds; (**f**) Meat rabbit lines. For single marker-based analysis: each dot represents a SNP. The blue line identifies the threshold value (99.95th percentile of the distribution). For window-based analyses: each dot represents a 350-kb genome window. The blue line identifies the threshold value (99.8th percentile of the distribution). Unassembled scaffolds are not included.


a) Albino breeds
b) Silver / greying breeds
c) Spotted breeds
d) Dwarf / small breeds
e) Giant breeds
f) Meat production
TYR
EDNRA
KITLG
LCORL/NCAPG
HOXD
COL11A1
GRID1
MSRA
HMGA2
CDH13
BMP5/COL21A1
SLC30A10
MRPL33
PTPN2
CCSER1
F_{ST}
Chromosome

Tables

Table 1 Details on the analysed breeds and animals

Breed	Acronym	Number of animals	Coat colour	Body size	Other information and characteristics
Meat breeds					
Italian Silver	ISI	20	Non-agouti black at birth and then silver	Medium	Derived from Champagne d'Argent
Italian Spotted	ISP	93	Himalayan coat colour pattern (white albino with black spots at the extremities)	Medium	Derived from Californian
Italian White	IW	256	White albino	Medium	Derived from New Zealand White
Fancy breeds					
Belgian Hare	BH	24	Brownish/dark red	Medium	
Burgundy Fawn	BF	6	Red/fawn	Medium	Genotype *e*/*e* at the *Extension* locus
Champagne d'Argent	CdA	19	Non-agouti black at birth and then silver	Medium	
Checkered Giant	CG	79	White with black spots	Giant	Heterozygous at the English spotted locus (*En/en* genotype)
Coloured Dwarf	CD	20	Various colours	Dwarf	
Dwarf Lop	DL	20	Various colours	Dwarf	Lopped ears
Ermine	ER	20	Greysh/cream	Small	
Giant Grey	GG	27	Grey/wild type	Large	
Giant White	GW	20	White albino	Large	
Rex	RE	19	Various colours	Medium	
Rhinelander	RH	28	Tricolour (white with black and red spots)	Medium	Heterozygous at the English spotted locus (*En*/*en* genotype). Genotype e^J/e^J at the *Extension* locus
Thuringian	TH	9	Pale red with dark shades	Medium	Genotype *e*/*e* at the *Extension* locus

Table 2 Overview of the investigated rabbit breeds and related population parameters

Breed	MAF (sd)	H_o (sd)	H_e (sd)	F_{IS} (sd)	M1-F_{ST} (sd)	M2-F_{ST} (sd)	N_e
Meat breeds							
Italian Silver	0.23 (0.15)	0.35 (0.19)	0.34 (0.16)	-0.03	0.14 (0.08)	0.23 (0.22)	35
Italian Spotted	0.23 (0.15)	0.34 (0.18)	0.33 (0.17)	-0.01	0.12 (0.08)	0.23 (0.22)	43
Italian White	0.27 (0.15)	0.36 (0.16)	0.36 (0.15)	-0.01	0.09 (0.07)	0.21 (0.20)	94
Fancy breeds							
Belgian Hare	0.20 (0.16)	0.34 (0.22)	0.32 (0.19)	-0.06	0.20 (0.13)	0.28 (0.26)	46
Burgundy Fawn	0.18 (0.16)	0.41 (0.26)	0.35 (0.19)	-0.15	0.23 (0.13)	0.29 (0.27)	17
Champagne d'Argent	0.19 (0.17)	0.33 (0.21)	0.32 (0.19)	-0.01	0.21 (0.13)	0.27 (0.26)	34
Checkered Giant	0.24 (0.16)	0.34 (0.17)	0.33 (0.16)	-0.01	0.13 (0.09)	0.22 (0.21)	86
Coloured Dwarf	0.25 (0.17)	0.29 (0.16)	0.35 (0.15)	0.17	0.14 (0.09)	0.22 (0.21)	54
Dwarf Lop	0.26 (0.16)	0.31 (0.17)	0.36 (0.16)	0.12	0.14 (0.09)	0.22 (0.21)	50
Ermine	0.18 (0.16)	0.28 (0.19)	0.31 (0.18)	0.09	0.23 (0.13)	0.28 (0.27)	33
Giant Grey	0.24 (0.15)	0.35 (0.19)	0.35 (0.16)	-0.01	0.13 (0.09)	0.22 (0.21)	63
Giant White	0.24 (0.16)	0.35 (0.20)	0.34 (0.17)	-0.02	0.14 (0.09)	0.22 (0.21)	43
Rex	0.25 (0.15)	0.32 (0.19)	0.35 (0.16)	0.08	0.14 (0.09)	0.22 (0.21)	40
Rhinelander	0.22 (0.17)	0.35 (0.21)	0.33 (0.18)	-0.05	0.17 (0.11)	0.26 (0.24)	42
Thuringian	0.21 (0.16)	0.40 (0.25)	0.35 (0.19)	-0.13	0.20 (0.12)	0.26 (0.25)	24

MAF: average minor allele frequency (standard deviation).
H_o: observed Heterozygosity (standard deviation).
H_e: expected Heterozygosity (standard deviation).
F_{IS}: inbreeding coefficient individual (I) relative to the subpopulation (S), (sd: standard deviation).
A M1-F_{ST}: average F_{ST} value computed by using Method 1 (standard deviation).
A M2-F_{ST}: average F_{ST} value computed by using Method 2 (standard deviation).
N_e: effective population size at the most recent generation (generation 13).

Table 3 The most important genomic regions and candidate genes derived from the window-based single-breed F_{ST} analyses

Traits	OCU: position[a]	Candidate gene	Breed (method: F_{ST} value)[b]	Previous studies[c]
Coat colour and structure	1:127563000-127668085	*TYR***	Burgundy Fawn (M1: 0.79; M2*: 0.62), Italian White (M1: 0.49; M2*: 0.50), Italian Spotted (M1: 0.54)	[15, 16]
	4:5435027-5439803	*ASIP***	Giant Grey (M1*: 0.49; M2: 0.53), Belgian Hare (M2*: 0.63)	[13]
	8:79700292-79724918	*EDNRB***	Rhinelander (M1: 0.64; M2: 0.63)	[79]
	9:37191805-37434390	*MITF*	Giant White (M1*: 0.44; M2: 0.54)	-
	14: 80045788-80094927	*LIPH***	Rex (M1*: 0.45; M2: 0.56)	[18]
	15: 17113617-17196501	*EDNRA***	Italian White (M1: 0.54; M2: 0.60), Thuringian (M2: 0.71)	-
	17:78036010-78397189	*OCA2***	Checkered Giant (M1: 0.62; M2: 0.57)	-
	Un0267:152309-153232	*MC1R*	Rhinelander (M2: 0.60)	[12]
Body size	2:8357629-8403513, 2:8404807-8620864	*NCAPG**, *LCORL***	Dwarf Lop (M2: 0.62), Ermine (M2: 0.72)	[19]
	13:36771716-36779306	*MEX3A***	Coloured Dwarf (M1: 0.69; M2: 0.61), Ermine (M1*: 0.70; M2*: 0.61)	-
	13:36843459-36895416	*ARHGEF2***	Coloured Dwarf (M1: 0.69; M2: 0.61), Ermine (M1*: 0.70; M2*: 0.61)	-
	13:37616687-37631494	*DCST1***	Coloured Dwarf (M1: 0.69; M2:0.61), Ermine (M1*: 0.70; M2*: 0.61)	-
	13:37644467-37654971	*ZBTB7B***	Coloured Dwarf (M1: 0.69; M2: 0.61), Ermine (M1*: 0.70; M2*: 0.61)	-
	13: 38651582-38702056	*GATAD2B***	Coloured Dwarf (M1: 0.55; M2: 0.55)	-
	13:61054286-61261797	*COL11A1*	Coloured Dwarf (M1*: 0.44), Ermine (M1*: 0.61)	[19]
	18: 66434580-66630682	*GRK5***	Dwarf Lop (M2: 0.63)	-
	Un0251:72110-94586	*COL2A1*	Dwarf Lop (M2: 0.57)	-

Supplementary information for this table is in Table S17 (see Additional file 1 Table S17)

[a]Position of the candidate gene, in basepairs, on the *O. cuniculus* reference genome (OryCun2.0). OCU = *Oryctolagus cuniculus* chromosome.

[b]Method used in the single breed approach: M1 or M2. F_{ST} value of the window including the candidate gene.

[c]Candidate genes previously reported in other studies in rabbit.

*Top 70 windows; all other F_{ST} results are from the top 14 windows containing the reported candidate gene; **Relevant genes have been identified based on F_{ST} window-based analysis and single marker-based analysis (see Additional file 1 Table S15 and S16).

F_{ST} values from both methods were reported (M1 and M2); in case for which the relevant identified gene was based only on one method (M1 or M2), the method is indicated.

Table 4 The most important genomic regions and candidate genes derived from the window-based F_{ST} analyses of groups of breeds

Traits	Comparison	OCU:poition[a]	Candidate gene	F_{ST}	Previous studies[b]
Albino breeds	(Italian white + Giant White) *vs* all other breeds	1:127563000-127668085	*TYR***	0.52	[15, 16]
		15: 17113617-17196501	*EDNRA***	0.58	-
Silver/graying of coat	(Italian Silver + Champagne d'Argent) *vs* all other breeds	Un513:25550-60684	*ADNP2***	0.60	-
		Un0329:192475-292759	*NFATC1*	0.54	-
Checkered/spotted phenotype	(Checkered Giant + Rhinelander) *vs* all other breeds	4: 67503403-67541344	*KITLG*	0.51	-
Dwarf/small body size	(Coloured Dwarf + Dwarf Lop + Ermine) *vs* all other breeds	2:8357629-8403513, 2:8404807-8620864	*LCORL***, *NCAPG***	0.50	[19]
		18:7460895-8191494	*GRID1***	0.59	-
		Un0030:1681671-2068484	*NTRK2***	0.54	-
		Un0030:348764-635797	*FRMD3***	0.57	-
		7:115660028-115662696	*HOXD*	0.40*	[19]
		13:61054286-61261797	*COL11A1*	0.42*	[19]
Large body size	(Checkered Giant + Giant Grey + Giant White) *vs* all other breeds	12:45671510-45793930; 12:45977386-46101862	*BMP5***, *COL21A1***	0.64	-
		16:82626965-83800640, 16:53149315-53161403	*CDH13***, *SLC30A10*	0.57	-
		2:39527266-39938880	*MSRA***	0.53	-
		4:44715759-44848250	*HMGA2***	0.40*	[19]
Meat production	Meat rabbit lines (Italian Silver + Italian Spotted + Italian White) *vs* all other breeds	2:158326199-158338803	*MRPL33***	0.57	-
		15:58389801-59653895	*CCSER1***	0.52	-
		9:49249318-49360593	*PTPN2***	0.51	-
		Un0030:1681671-2068484	*NTRK2*	0.49	-

[a]Position of the candidate gene, in bp, on the *O. cuniculus* reference genome (OryCun2.0). OCU = *Oryctolagus cuniculus* chromosome.

[b]Candidate genes previously reported in other studies in rabbit.

*Top 70 windows; all other F_{ST} results are from the top 14 windows containing the reported candidate gene; **Relevant genes have been identified based on F_{ST} window-based analysis and single marker-based analysis (see Additional file 1 Table S21).

Chapter 4: Comparative analysis of genomic inbreeding parameters and runs of homozygosity islands in several fancy and meat rabbit breeds

Mohamad Ballan[1], Giuseppina Schiavo[1], Samuele Bovo[1], Michele Schiavitto[2], Riccardo Negrini[3], Andrea Frabetti[4], Daniela Fornasini[4] and Luca Fontanesi[1]*

[1]Division of Animal Sciences, Department of Agricultural and Food Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy

[2]Associazione Nazionale Coniglicoltori Italiani (ANCI), Contrada Giancola snc, 71030 Volturara Appula (FG), Italy

[3]Associazione Italiana Allevatori, Via G. Tomassetti 9, 00161 Roma, Italy

[4]Gruppo Martini spa, Via Emilia n. 2614, 47020 Budrio di Longiano (FC), Italy

Published in Journal of Animal Genetics. 08 September 2022. Doi: https://doi.org/10.1111/age.13264

Summary

Runs of homozygosity (ROH) are defined as long stretches of DNA homozygous at each polymorphic position. The proportion of genome covered by ROH and their length are indicators of the level and origin of inbreeding. In this study, we analysed SNP chip datasets (obtained using the Axiom OrcunSNP Array) of a total of 702 rabbits from 12 fancy breeds and 4 meat breeds to identify ROH with different methods and calculate several genomic inbreeding parameters. The highest average number of ROH per animal was detected in Belgian Hare (n. ~150) and the lowest in Italian Silver (n. ~106). The average length of ROH ranged from 4.001 ± 0.556 Mb in Italian White to 6.268 ± 1.355 Mb in Ermine. The same two breeds had the lowest (427.9 ± 86.4 Mb, Italian White) and the highest (921.3 ± 179.8 Mb, Ermine) average values of the sum of all ROH

segments. Most fancy breeds had a higher level of genomic inbreeding (as defined by ROH) than meat breeds. Several ROH islands contains genes involved in body size, body length, pigmentation processes, carcass traits, growth and reproduction traits (e.g.: *AOX1, GPX5, IFRD1, ITGB8, NELL1, NR3C1, OCA2, TRIB1, TRIB2*). Genomic inbreeding parameters can be useful to overcome the lack of information in the management of rabbit genetic resources. ROH provided information to understand, to some extent, the genetic history of rabbit breeds and to identify signatures of selection in the rabbit genome.

Keywords: Genetic variability; *Oryctolagus cuniculus*; ROH; Signature of selection; SNP

Introduction

The domestication process of the European rabbit (*Oryctolagus cuniculus*), usually referred as domestic rabbit or simply rabbit, relied on wild populations that colonized the South of France (Zeuner 1963). The process probably started quite recently, in a period that spanned from the high middle age to the XV-XVI centuries. In the early stages, domestication of this species might be mainly associated with the activities of French monasteries and castles (Zeuner 1963; Callou 2002; Zeder 2012). Subsequently, it continued with more undefined trajectories that followed the spread of the rabbit in the North-Central Europe, till the constitution of some breeds (reviewed in Fontanesi 2021a; Fontanesi *et al.* 2021b). The domesticated rabbit genetic pool became slightly different from the wild counterparts in terms of allele frequencies at many regulatory sites affecting brain and neuronal development, with potential impact on the behaviour of the animals, that, in this way, could be more easily handled and bred in captivity (Carneiro *et al.* 2014). Then, the human driven artificial selection, that led to the constitution of the modern rabbit breeds, mainly worked on exterior traits (e.g. coat colours, body size, ear length). The result was that a broad phenotypic diversity distinguished many breeds valued by fancy breeders who continued to create additional breeds, lines or strains by introgressing specific features and creating new combinations of traits (Boucher *et al.* 2021; Fontanesi 2021a,b). Most fancy breeds are named according to their colouration or other peculiar exterior features and specific standards have been defined by breeders' organizations or societies, that are present in several countries (Whitman 2004; Boucher *et al.* 2021). Some of these breeds are recognized by more than one breeders' societies or associations, other breeds are recognized only at the national level. More recently, specialized meat lines have been also constituted by selecting the animals for performance and production traits.

Both fancy breeds (with their close or semi-close national organizations and structures, with few exchanges of animals with other countries) and meat lines (where selection is usually carried out within close nuclei) can be genetically considered small populations. Therefore, to properly manage these populations, it is important to monitor their level of inbreeding. An increased level of autozygosity leads to inbreeding depression and the occurrence of deleterious recessive alleles in homozygous state, which mainly impairs reproductive performances that are usually quite poor, particularly in fancy rabbit purebreds (Boucher *et al.* 2021).

In diploid organisms, inbreeding (traditionally indicated with F_{PED} when based on pedigree information) can be defined as the probability that, at a randomly selected locus, the two alleles derived from the maternal and the paternal sides are identical by descent (Wright 1922). This definition of F_{PED} can be extended as the proportion of all loci of an individual's genome that is identical by descent. In a population, the level of inbreeding is estimated by averaging all F_{PED} individual values.

In rabbits, where pedigree recording systems and DNA based methods to control parentage are not well established, reliability of F_{PED} values is in general quite low. Therefore, despite that a few assumptions used to estimate F_{PED} can be considered only as approximations needed by the method of calculation (Leutenegger *et al.* 2003; Wang 2016; Knief *et al.* 2017), in many cases F_{PED} could not be calculated at all. Genomic tools now available also in *O. cuniculus* [a reference genome and a single nucleotide polymorphism (SNP) array; Carneiro *et al.* 2014; Fontanesi *et al.* 2021a] can provide some alternatives, which might overcome the limits that F_{PED} have and that are quite relevant in many rabbit populations.

As already reported in several other species, the level of autozygosity of an animal can be estimated interrogating the genotype status at thousands of polymorphic sites obtained by

genotyping SNPs covering the whole genome (e.g. Kristensen *et al.* 2010). A few inbreeding related parameters can be calculated using genome-wide SNP data (Leutenegger *et al.* 2003; Van Raden 2008; Kardos *et al.* 2015; Schiavo *et al.* 2020b). Among these parameters, runs of homozygosity (ROH), defined as continuous chromosome stretches in which all loci have a homozygous genotype (Gibson *et al.* 2006), if used to sum up the proportion of the genome in autozygosity state, can provide a quite precise measure of genomic inbreeding of an individual animal (F_{ROH}; Kardos *et al.* 2015; Peripolli *et al.* 2017). At the population level, some other characteristics of ROH (the patterns of ROH distribution across the chromosomes, the average length of ROH and the average proportion of the genome covered by ROH) can provide some indications to infer the genetic history of the populations (Ceballos *et al.* 2018). ROH can also be useful to identify signatures of selection: a high frequency of ROH in certain chromosome regions (defined as ROH islands or ROH hotspots) highlights reduced haplotype variability that spans loci under artificial selection or natural selection, as already reported in several livestock species (Purfield *et al.* 2017; Bertolini *et al.* 2018; Mastrangelo *et al.* 2018; Peripolli *et al.* 2018; Grilz-Seger *et al.* 2019; Gorssen *et al.* 2020; Schiavo *et al.* 2020a, 2021).

In this study we analysed SNP array datasets from a total of 16 fancy and meat rabbit breeds to identify ROH and calculate several genomic inbreeding parameters. Then, we evaluated the occurrence of ROH islands in the genome of some of these breeds to identify putative signatures of selection that might be derived by different selection histories and structures of these rabbit genetic resources.

Materials and methods

Animals

Biological specimens (hair roots or buccal swaps) were sampled from a total 712 rabbits from four meat and 12 fancy rabbit breeds. Three meat breeds (Italian White, n. 256; Italian Spotted, n. 93; Italian Silver, n. 20) and all fancy breeds (Belgian Hare, n. 24; Burgundy Fawn, n. 6; Champagne d'Argent, n. 19; Checkered Giant, n. 79; Coloured Dwarf, n. 20; Dwarf Lop, n. 20; Ermine, n. 20; Giant Grey, n. 27; Giant White, n. 20; Rex, n. 19; Rhinelander, n. 28; and Thuringian, n. 9) were from the national Herd Book maintained by the Italian Rabbit Breeders Association (ANCI). One meat breed (n. 52) was from another albino white nucleus of Gruppo Martini spa selected for meat production, indicated thereafter Commercial Meat line. All breeds from the ANCI Herd Book had the standard breed characteristics. The description of the breeds is reported in Table S1. All animals included in this study were selected to avoid highly related individuals (no full- or half-sibs).

SNP genotyping and population genomic analyses

DNA extraction was carried out using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). DNA from each rabbit was then genotyped with the Affymetrix Axiom OrcunSNP Array (Thermo Fisher Scientific/Affymetrix Inc., Santa Clara, CA, USA). Quality control of the genotyping data was made with the Axiom Analysis Suite that led to discard low quality SNPs and with PLINK v.1.9 software (Chang *et al.* 2015) using the following filtering criteria: SNPs with more than 10% missing genotypes were excluded from the analysis; animals were discarded if their call rate was <0.90. We avoided filtering the dataset based on minor allele frequency (MAF) because this procedure could lead to the underestimation of the coverage in ROH (Meyermans *et al.* 2020). Finally, a total of 43529 SNPs and 702 animals remained for further analyses (Table S1).

Genotyping data were used to produce multidimensional scaling (MDS) plots with the first three dimensions computed with PLINK v.1.9 software (Chang *et al.* 2015). The effective population size (*Ne*) for recent and distant generations was calculated with default parameters using SNeP software (Barbato *et al.* 2015).

Runs of homozygosity and other genomic inbreeding measures

PLINK v.1.9 software (Chang *et al.* 2015) was used to identify ROH. No pruning based on linkage disequilibrium was performed to avoid biases that could be derived from this procedure (Marras *et al.* 2015; Meyermans *et al.* 2020). A minimum length of 1 Mb was chosen to detect ROH. This threshold may exclude short and common ROH determined by markers in linkage disequilibrium (Ferenčaković *et al.* 2013b; Marras *et al.* 2015). Four methods were used to identify ROH by varying and combining a few calling parameters: i) the minimum number of consecutive homozygous SNPs included in the ROH was 15 (Method 1 and Method 2) or 50 (Method 3 and Method 4); ii) the minimum length that constituted the ROH was 1 Mb; iii) the number of heterozygous SNPs that were allowed in the ROH was 0 (Method 1 and Method 3) or 2 (Method 2 and Method 4); iv) the minimum density of SNPs in a genome window was 1 SNP every 100 kb; v) the maximum gap between consecutive SNPs was 1000 kb. The parameters used for the four methods are summarised in Table S2. Identified ROH were then grouped into five size classes according to their physical length (Kirin *et al.* 2010; Ferenčaković *et al.* 2013a; Schiavo *et al.* 2020b, 2021): ROH1-2 (ROH ≥ 1Mb and < 2 Mb); ROH2-4 (ROH ≥ 2 Mb and < 4 Mb); ROH4-8 (ROH ≥ 4 Mb and <8 Mb); ROH8-16 (ROH ≥ 8 Mb and <16 Mb); ROH>16 (ROH ≥16 Mb). Other ROH parameters were also computed for each rabbit: 1) number of ROH (nROH); 2) the

average length of ROH (L_{ROH}, in Mb); 3) the sum of all ROH segments (S_{ROH}, in Mb). The average values of those parameters were also calculated for each breed.

Genomic inbreeding measures (F_{ROH}) were obtained as the proportion of the autosomal genome covered by ROH. According to multiple detected ROH classes with length >1 Mb classes, >4 Mb, >8 Mb and >16 Mb, the inbreeding coefficients F_{ROH1}, F_{ROH4}, F_{ROH8} and F_{ROH16} were obtained respectively. Moreover, the individually F_{ROH} values were averaged to calculate the final inbreeding coefficients values for each rabbit breed included in the study.

Furthermore, using PLINK software version 1.9 (Chang *et al.* 2015) with the ported functions of GCTA software v. 1.92 (Yang *et al.* 2011), other different inbreeding coefficients were also considered. These measures included: i) the genomic inbreeding coefficient F_{HOM} based on the difference between observed and expected number of homozygous genotypes; ii) the inbreeding coefficient based on the genomic relationship matrix F_{GRM} (Van Raden *et al.* 2011; this parameter is equivalent to F_{hat1}, even if scaled in a different way); iii) the estimate based on the variance-standardized relationship minus 1 (F_{hat1}); iv) the excess of homozygosity-based inbreeding estimate (F_{hat2}); v) the estimate based on correlation between uniting gametes (F_{hat3}). Finally, we estimated the Pearson correlation coefficients (*r*) between all pairs of inbreeding coefficients.

ROH islands and annotation of genome regions

First, the proportion of SNPs residing within a ROH was calculated for a given breed by counting the number of times a SNP appeared in a ROH within the given breed divided by the total number of genotyped rabbits of that breed. Then, a percentile threshold of the frequency of a SNP in ROH was calculated. The percentile-based threshold was defined considering the top 1% of

SNPs observed in a ROH in each breed. Adjacent SNPs, having a frequency of ROH occurrences over or equal the identified thresholds, and with a distance ≤ 1 Mbp between them, constituted ROH islands (Schiavo *et al.* 2021).

The identified ROH islands were annotated with the Bedtool v.2.17 (https://bedtools.readthedocs.io/) by retrieving the mapped genes form OryCun2.0 NCBI's GFF file. Functional enrichment analysis was carried out with Enrichr (Chen *et al.* 2013) via Fisher's exact test. Analyses was performed according to the following databases: GWAS catalog 2019, KEGG Human database 2019, MGI mammalian phenotype level dataset 2019 and the Biological process branch of Gene Ontology (GO). The analysis was executed for each breed specifically by using the set of genes mapped with ROH islands as input. Additionally, the statistically over-represented terms were considered if at least two input genes from two or more different ROH islands were involved and if the adjusted *P* value was lower than 0.25.

Results

Population genomic parameters

The tri-dimensional MDS-plot of Figure S1 graphically represented the genomic information and relationships of the rabbit breeds based on the SNP dataset. Rabbits belonging to the same breed were distinguishably clustered together. Two of the meat breeds (Italian White and the Commercial meat breed) were grouped separately among other breeds. Other breeds that share the same or similar morphological traits were grouped closely together. For example: Ermine rabbits (from a breed with animals of small size) were close to the Coloured Dwarf breed; and the three giant breeds (Checkered Giant, Giant Grey and Giant White) were grouped together. All other breeds could not be clearly resolved with this approach.

The estimated effective population size (N_e) of the 16 rabbit breeds is reported in Table S3. Italian White, Giant Grey and the Commercial Meat breeds had the highest values (at the 13th past generation: 94, 88 and 73, respectively). The lowest N_e was in Ermine, Thuringian and Burgundy Fawn breeds (34, 25 and 18, respectively). It is worth to note that Burgundy Fawn and Thuringian were the breeds with the lowest number of analysed rabbits (Table S1), which could have biased this parameter.

Runs of homozygosity in 16 rabbit breeds

We used different parameters to call ROH that were summarised in four methods (Table S2). Method 2, which used two relaxed parameters (a minimum number of 15 consecutive homozygous SNPs to call ROH and a maximum of 2 heterozygous SNPs in the homozygous SNP-windows), was empirically considered the most appropriate method in our case, as it can better overcome some of the current limitations of the genomic tools available in rabbits (Fontanesi *et al*. 2021b): the OryCu2.0 genome version that has the N50 length for the contigs equal to 64.65 kb; the genotyping platform for which the useful SNPs, after all filtering, were about 1/5 of the total number of genotyped SNPs. Therefore, all main ROH results were based on this method.

Table 1 provides an overview of the ROH identified in the investigated breeds, using Method 2, whereas a comparative analysis among breeds and methods is reported in Table S4. The average number of ROH per animal (nROH) ranged from 105.5 (but with a quite large SD: 32.0) in Italian Silver to 150.2 ± 13.2 in Belgian Hare. The average length of ROH (L_{ROH}) ranged from 4.001 ± 0.556 Mb (Italian White) to 6.268 ± 1.355 Mb (Ermine). The same two breeds had the lowest (427.9 ± 86.4 Mb, Italian White) and the highest (921.3 ± 179.8 Mb, Ermine) S_{ROH} values. The correlation for these three ROH statistics (i.e. nROH, L_{ROH} and S_{ROH}) determined with the four

different methods are reported in Tables S5, S6 and S7. Correlations for these parameters determined with Method 2 between the same parameters calculated with the other methods ranged from 0.882 to 0.996 (for nROH), from 0.773 to 0.998 (for L_{ROH}) and from 0.909 to 0.999 (for S_{ROH}). The highest correlations were between methods that used the same maximum number of allowed heterozygous SNPs included in the homozygous window (i.e. M2 *vs* M4 and M1 *vs* M3), whereas the minimum number of SNPs included in the window (15 or 50) did not have any relevant effects of the considered ROH parameters (Tables S5, S6 and S7).

Figure 1 represents the correlation plots between nROH and S_{ROH} over all individuals of each of the 16 rabbit breeds. Homogeneous correlation plots are evident in Champagne d'Argent, Coloured Dwarf, Commercial Meat and Giant Grey indicating that most animals within these breeds had similar ROH parameters (nROH, L_{ROH} and S_{ROH}). In contrast, a heterogeneous distribution was observed in Checkered Giant, Dwarf Lop, Italian Silver, Italian Spotted and Rex breeds, whereas all other breeds had intermediate patters (excluding the breeds for which a small number of animals was genotyped).

The proportion of ROH of different class length for the 16 breeds is shown in Figure 2. Table S8 reports the detailed numbers. Ermine, Coloured Dwarf and Italian Silver had the highest proportion of long ROH (>16 Mb) about (7%, 6% and 5%, respectively). Short ROH classes (1-4 Mb) were more frequent in the three giant breeds, in the Commercial Meat and Italian White breeds (Table S8).

The largest ROH in term of length was identified in Ermine and Coloured Dwarf (87.4 Mb on OCU3). These two breeds also included animals with the largest S_{ROH} values in the whole dataset: a Coloured Dwarf rabbit had a S_{ROH} of 1304.3 Mb and an Ermine rabbit had a S_{ROH} of 1226.5 Mb, which indicate that almost half of their genomes were covered by ROH. The same

Ermine rabbit had also the largest number of ROH that we identified in this study (n. = 192). Table S9 reports the minimum and maximum values observed for nROH, L_{ROH} and S_{ROH} in all breeds.

Genomic inbreeding parameters

The mean and standard deviation of genomic inbreeding parameters calculated using different classes of ROH length (F_{ROH1} to F_{ROH16}) are reported in Table 2 (for a complete evaluation, Table S10 includes results obtained with all four Methods). With Method 2, Italian White and Commercial Meat rabbits had the lowest F_{ROH1} values (0.219 and 0.255, respectively). The highest F_{ROH1} values were obtained in Ermine, Burgundy Fawn and Champagne d'Argent (0.472, 0.400 and 0.385, respectively). Considering the inbreeding parameters based on only medium-long ROH (F_{ROH4}, F_{ROH8} and F_{ROH16}), the values decreased in all breeds, as expected, but with remarkable differences. For example, in the Commercial Meat rabbits, F_{ROH16} dropped 14 times whereas in Ermine, this parameter dropped only 3.9 times compared to the corresponding F_{ROH1} values. The distribution of the F_{ROH} values of the different length classes obtained with Method 2 is reported in the boxplots of Fig. 3.

Other five genomic inbreeding parameters have been calculated (F_{hat1}, F_{hat2}, F_{hat3}, F_{GRM} and F_{HOM}). The average values of F_{hat1}, F_{hat2}, F_{hat3} and F_{HOM} were negative in all breeds except two breeds for F_{hat1} (Coloured Dwarf and Dwarf Lop) or four breeds for F_{hat2}, F_{hat3} and F_{HOM} (Coloured Dwarf, Dwarf Lop, Ermine and Rex), even if with large standard deviation (Table S11). These results could be interpreted, to some extent, that in the breeds with positive values, rabbits were on average more related to each other than what happened for the rabbits of the breeds with negative values, even if there was a large within-breed variability for all parameters. Distribution

plots of all these other five inbreeding parameters in the 16 rabbits breeds are shown in Figs S2 & S3.

Table S12 reports the correlation between all considered genomic inbreeding measures calculated within each breed. The correlation between the F_{ROH} parameters based on different ROH length was high in all breeds. Correlation values ranged from 0.997 in Champagne d'Argent (F_{ROH1} *vs* F_{ROH4}) to 0.582 in Italian White (F_{ROH1} *vs* F_{ROH16}). Values were higher between close length classes than between distant close classes: that means that correlations in all breeds were higher if, for example, F_{ROH1} and F_{ROH4} were considered in the pairwise analysis than if F_{ROH1} was compared with F_{ROH8} or F_{ROH16}. This was expected considering the distribution of ROH classes (Fig. 3) and the progressive reduction of longer ROH in all breeds. However, this drop of correlation was less relevant in Champagne d'Argent, Coloured Dwarf and Ermine where all correlations were > 0.90. The correlation between F_{ROH} parameters and all other genomic inbreeding parameters were high and consistent over all breeds only with the F_{HOM} values (Table S12). For example, correlation between F_{ROH1} and F_{HOM} ranged from 0.974 (Giant Grey) to 0.783 (Dwarf Lop) and correlation between F_{ROH16} and F_{HOM} ranged from 0.939 (Champagne d'Argent) to 0.597 (Belgian Hare). Correlations between the other parameters were not always consistent across breeds (excluding F_{hat1} and F_{GRM}, which are equivalent parameters), with a large range that in some cases spanned extreme values (Table S12).

Runs of homozygosity islands

As the interpretation of the results in terms of potential signature of selection is much more reliable when many animals are genotyped (Ceballos *et al.* 2018), we considered in more details the information derived from the four breeds (Checkered Giant, Commercial Meat, Italian Spotted

and Italian White) for which more than 50 animals were genotyped (summarizing information for all 16 breeds are presented in Tables S13 and S14). Manhattan plots reporting ROH islands in these four breeds are shown in Fig. 4. In these breeds, ROH islands were identified in a total of 11 autosomes, which covered about 10-20 Mb of the rabbit genome (10.5 Mb in Checkered Giant, 10.5 Mb in Commercial Meat, 17.7 Mb in Italian Spotted and 19.7 Mb in Italian White). Considering overlapping ROH islands between breeds, a total of 22 independent ROH islands were detected (Table 3). A few ROH islands were identified in all four breeds (on OCU3, position ~140.60-142.31 Mb) or in three breeds (on OCU7 and OCU15) or in two breeds (on OCU7, two in OCU12 and on OCU15). Among the list of the annotated genes included in the ROH islands, some interesting candidate genes, which might provide some potential functional relationships with phenotypic features of the corresponding breeds, could be identified. Several ROH islands contains genes involved in body size, body length, carcass traits, growth and reproduction traits as previously determined in humans or in other livestock species, including the rabbit (Table 3). Considering that most of the ROH islands were identified in meat breeds, it is interesting to note that several genes included in these regions have been already reported to be associated with production traits mainly relevant in meat species (Table 3). Some of these genes are also associated to relevant human traits. For example, the ROH island on OCU3 identified in all four breeds encompasses eight different genes, including the *tribbles pseudokinase 1* (*TRIB1*) gene. TRIB1 is a signalling regulator protein involved in the activation and suppression of the various interacting signalling pathways. Mutations in the human gene are strongly associated with several fat deposition traits, serum metabolite contents, including triglyceride and cholesterol levels. Another member of this gene family, *tribbles pseudokinase 2* (*TRIB2*) gene, is included in another ROH located on OCU2 and identified in Italian Spotted. The ROH island on OCU17 identified in

Checkered Giant contains the *oculocutaneous albinism type 2* (*OCA2*) gene, which is involved in the pigmentation processes (e.g. Donnelly *et al.* 2012).

The genome enrichment analysis for the genes within ROH islands showed over-representation of several GO terms (Table S15). Considering the four breeds mentioned above, Checkered Giant had three enriched terms (atrial fibrillation, basal cell carcinoma, body mass index) and the Commercial Meat line had two enriched terms (brest cancer; systolic blood pressure) whereas no enriched term was identified in Italian White and Italian Spotted. Among the enriched terms identified in the other 12 breeds, it was interesting to note that in Dwarf Lop an enriched term was "height" (Table S15).

Discussion

Fancy rabbit breeds, with their national breeders' associations and close structures, and meat rabbit lines, with their close nuclei in which a small number of animals are performance tested, are interesting examples of livestock populations where inbreeding should be carefully managed. Empirical experience of practitioners working with rabbits indicates that pedigree recording is usually not very precise (mainly in fancy breeds) which calls for alternative approaches to obtain good estimation of inbreeding. As already demonstrated in many livestock species, genomic analyses could overcome the low accuracy of the pedigree registrations and the biased assumptions that underline the pedigree-based estimations of inbreeding (e.g. Howard *et al.* 2017; Schiavo *et al.* 2020). In rabbit, however, the routine application of SNP genotyping is still in its infancy, mainly due to the high genotyping cost compared to the value of the animals and to a lower level of organization of the breeder/breeding industries than what has been already achieved in other livestock species. Applications of genome-wide analyses in domestic rabbit populations has been

mainly carried out to understand the effect of selection pressures over generations and to identify genes affecting exterior traits and QTL for economically relevant traits, with limited use of the genomic information for other purposes (e.g. Sosa-Madrid *et al.* 2020a, 2020b, 2020c; Laghouaouta *et al.* 2020; Bovo *et al.* 2021; Carneiro *et al.* 2021; Liu *et al.* 2021).

In this study we genotyped, with a SNP array, rabbits belonging to 16 breeds and populations and used this information to estimate genomic inbreeding parameters, including different ROH measures (mainly used for the comparative analyses), and to detect signatures of selection defined by ROH hotspots.

Several parameters can be adjusted to call ROH (e.g. number of missed genotypes, number of heterozygous SNP allowed in ROH, minimum size of a ROH and so on), which can affect the final outputs. We tested some of them as there is no general rule established in this context (Peripolli *et al.* 2017) and we needed to tune parameters based on the characteristics of the genotype dataset and of the reference genome OryCun2.0, which has a limited N50. By modifying the number of allowed heterozygous SNPs, the parameters that was mainly affected was the average length of the ROH, as expected. The methods that allowed up to two heterozygous SNPs (Method 2 and 4) had larger average length of ROH than the methods that did not allow any heterozygous SNPs (Methods 1 and 3). The correlations of the two groups of methods for different ROH measures was however very high (from 0.761 to 0.774, for the average length of ROH; from 0.882 to 0.886 for nROH; from 0.909 to 0.921 for S_{ROH}; Tables S5, S6 and S7). Among these parameters S_{ROH} is directly related to F_{ROH} measures, suggesting that the general overview obtained by different methods is similar in terms of interpretation of the results for the analysed breeds.

The most critical breed in terms of inbreeding level of the population was Ermine (a rare breed) that had the highest F_{ROH} values and a small Ne. Among the other fancy breeds, Checkered Giant had the lowest F_{ROH} values, indicating a low level of inbreeding. This reflects the fact that rabbits with the Checkered Giant breed-specific spotted patterns are actually heterozygous animals at the *English spotting* locus (with genotype *En*/*en*; Fontanesi *et al.* 2014). These animals can be obtained by cross-breeding programs, which might increase the level of heterozygosity at many other loci. All three meat breeds had, in general, lower F_{ROH} values than most of the remaining fancy breeds, suggesting that their nuclei are well managed using pedigree recording information to monitor inbreeding.

Among the other genomic inbreeding parameters that we considered in this study (F_{hat1}, F_{hat2}, F_{hat3}, F_{GRM} and F_{HOM}) only F_{HOM} was consistent across breeds and with high to moderate correlation with F_{ROH} measures. This general picture is consistent to what we already reported in pigs where F_{HOM} produced similar results of F_{ROH} (Schiavo *et al.* 2020).

Among the inbreeding genomic measures that we calculated, ROH characteristics can be interpreted to infer the genetic structure of the analysed breeds (Ceballos *et al.* 2018). Starting from the basic ROH parameters, it is possible to read interesting elements. Standard deviation of ROH measures (Table 1) was not very high compared to what we reported in a similar study for different autochthonous and cosmopolitan pig breeds in Europe (Schiavo *et al.* 2021). This parameter, together with the results of MDS-plot (Figure S1), might suggest that few substructures or subpopulations were sampled in the analysed breeds. Other population genomic statistics should be estimated to better evaluate these aspects. It would be also interesting to genotype a larger number of animals per breed, including rabbits sampled in different countries but of the same breeds to better evaluate these aspects.

The genetic history of the investigated rabbit breeds can be also inferred, to some extent, from other ROH measures (Ceballos *et al.* 2018). Following the assumption that recent inbreeding usually generates long ROH whereas short ROH have a common ancestral origin (Kirin *et al.* 2010; Ceballos *et al.* 2018), it seems that most rabbit breeds included in this study could be mainly ascribed to the second condition. Their genome was covered by many ROH per animal (on average, from about 106 in Italian Silver to about 150 in Belgian Hare) but with quite short L_{ROH} (on average, from ~4 Mb in Italian White to ~6.3 Mb in Ermine). This overview does not change if the method used to call ROH is changed (we used four methods; see Table S4). This general picture in these rabbit breeds is however the opposite to what we already observed in a study in which we investigated ROH in European pig breeds, where nROH was much lower on average per breed with a larger average size of L_{ROH}, mainly due to recent inbreeding events and bottlenecks (Schiavo *et al.* 2021). Therefore, it seems that in most of these rabbit breeds, within breed identical-by-descent chromosome segments might be shared by old ancestors. This interpretation is in agreement with the need, in many fancy breeds, to keep fixed originally defined breed-specific features that have, in several cases, a monogenic or oligogenic determinism and that have been considered the original and basic elements of the standard of the breeds (Fontanesi 2021).

Another study used ROH to estimate genomic inbreeding parameters in a rabbit line established in the 1980' and selected for reproduction and growth traits (Rodríguez-Ramilo *et al.* 2020). Despite ROH were calculated in different ways, the F_{ROH}, estimated by summing up the values obtained from size ranges, is similar to what we obtained in the meat breeds. This might suggest again that fancy rabbit breeds should be considered apart from other rabbit lines where inbreeding is usually avoided to maintain high production performances.

Runs of homozygosity were also investigated in rabbits by Casto-Rebollo *et al.* (2021) for another purpose. These authors used ROH to identify signatures of divergent selection in a rabbit population from which divergent lines for environmental variance of litter size were established. In that study, ROH were more effective than other population genomic parameters to identify signatures of selection. We also analysed ROH to identify hotspots of selection in the 16 investigated breeds. Due to the low number of animals genotyped for a few breeds, we mainly focused our attention on the breeds for which more than 50 animals were investigated. Some of the ROH islands detected in Checkered Giant, Commercial Meat, Italian Spotted and Italian White breeds contained genes involved in growth, body size and several other important traits in meat species, including carcass, meat and fat deposition traits, growth rate, feed efficiency, stress sensitivity and reproduction traits. It was interesting to mention the ROH island encompassing the *OCA2* genome region that was identified in Checkered Giant (also known as Papillion). This gene encodes for the homolog of the mouse p (pink-eyed dilution) that is believed to be a melanocyte-specific transporter with an essential role in normal pigmentation. The role of this gene is however not completely well defined yet. In another study that involved Checkered Giant rabbits (Ballan *et al.* 2022), using F_{ST} parameters we confirmed the signature of selection in this gene region. Therefore, it is quite remarkable that two independent types of analyses (i.e. ROH islands and F_{ST}) pointed out this region in this breed that is mainly defined by a peculiar spotted pattern derived by an heterozygous state at the *English spotting* locus, which is associated with alleles at the *KIT proto-oncogene, receptor tyrosine kinase* (*KIT*) gene (Fontanesi *et al.* 2014). Potential interactions between the *OCA2* and *KIT* gene products, not yet well established, could be needed to determine the Checkered Giant spotted design.

Genomic analyses in the rabbit can be important for several reasons that span from practical aspects, including the management of small populations with genomic inbreeding parameters, to the exploration of the large genetic diversity that is present across breeds and within breeds and that can be used to identify the genetic architecture underlining phenotype diversity. Runs of homozygosity can complement other methods that are used to interpret the genetic history of livestock populations and to detect signatures of selection in rabbit breeds, exploiting unique genetic resources available from fancy breeders and breeding industries.

Acknowledgements

This study was funded by the PSRN (Programma di Sviluppo Rurale Nazionale) Cun-Fu and Cun-Fu 2 projects (co-funded by the European Agricultural Fund for Rural Development of the European Union and by the Italian Ministry of Agriculture, Food and Forestry - MiPAAF), by the University of Bologna RFO 2019-2021 programs and by Gruppo Martini spa (project Cun-Genomics).

Conflict of interest

The authors declare they do not have any competing interests.

Data availability

Genotyping data can be shared after the signature of an agreement on their use with the University of Bologna and ANCI. Please address all requests to luca.fontanesi@unibo.it.

References

An B., Xia J., Chang T., Wang X., Miao J., Xu L., Zhang L., Gao X., Chen Y., Li J. & Gao H. (2018) Genome-wide association study identifies loci and candidate genes for internal organ weights in Simmental beef cattle. *Physiological Genomics* **50**, 523–31.

Ballan M., Bovo S., Schiavo G., Schiavitto M., Negrini R. & Fontanesi L. (2022) Genomic diversity and signatures of selection in meat and fancy rabbit breeds assessed using high density SNP information. *Genetics Selection Evolution* **54**, 3.

Barbato M., Orozco-terWengel P., Tapio M. & Bruford M.W. (2015) SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* **6**, 109.

Barranco I., Tvarijonaviciute A., Perez-Patino C., Vicente-Carrillo A., Parrilla I., Ceron J.J., Martinez E.A., Rodriguez-Martinez H. & Roca J. (2016) Glutathione peroxidase 5 is expressed by the entire pig male genital tract and once in the seminal plasma contributes to sperm survival and in vivo fertility. *PLoS One* **11**, e0162958.

Bertolini F., Cardoso T.F., Marras G., Nicolazzi E.L., Rothschild M.F. & Amills M. (2018) Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution* **50**, 59.

Brunes L.C., Baldi F., Lopes F.B., Lôbo R.B., Espigolan R., Costa M.F., Stafuzza N.B. & Magnabosco C.U. (2021) Weighted single-step genome-wide association study and pathway analyses for feed efficiency traits in Nellore cattle. *Journal of Animal Breeding and Genetics* **138**, 23–44.

Bovo S., Schiavo G., Utzeri V.J., Ribani A., Schiavitto M., Buttazzoni L., Negrini R. & Fontanesi L. (2021) A genome-wide association study for the number of teats in European rabbits (*Oryctolagus cuniculus*) identifies several candidate genes affecting this trait. *Animal Genetics* **52**, 237–43.

Callou C. (2002) De la garenne au clapier. Étude archéozoologique du Lapin en Europe occidentale. Mémoires do Muséum National d'Histoire Naturelle. 189. Paris, France.

Carneiro M., Vieillard J., Andrade P., Boucher S., Afonso S., Blanco-Aguiar J.A., Santos N., Branco J., Esteves P.J., Ferrand N., Kullander K. & Andersson L. (2021) A loss-of-function mutation in RORB disrupts saltatorial locomotion in rabbits. *PLoS Genetics* **17**, e1009429.

Carneiro M., Rubin C.J., Di Palma F., Albert F.W., Alföldi J., Martinez Barrio A., *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–9.

Casto-Rebollo C., Argente M.J., García M.L., Blasco A. & Ibáñez-Escriche N. (2021) Selection for environmental variance of litter size in rabbits involves genes in pathways controlling animal resilience. *Genetics Selection Evolution* **53**, 59.

Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M. & Wilson J.F. (2018) Runs of homozygosity: Windows into population history and trait architecture. *Nature Review Genetics* **19**, 220–34.

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M. & Lee J.J. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7.

Chang T., Xia J., Xu L., Wang X., Zhu B., Zhang L., Gao X., Chen Y., Li J. & Gao H. (2018) A genome-wide association study suggests several novel candidate genes for carcass traits in Chinese Simmental beef cattle. *Animal Genetics* **49**, 312–6.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V., Clark N.R. & Ma'ayan A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128.

de Lima A.O., Koltes J.E., Diniz W.J., de Oliveira P.S., Cesar A.S., Tizioto P.C., Afonso J., de Souza M.M., Petrini J., Rocha M.I.P., Cardoso T.F., Neto A.Z, Coutinho L.L., Mourão G.B. & Regitano L.C. (2020) Potential biomarkers for feed efficiency-related traits in nelore cattle identified by co-expression network and integrative genomics analyses. *Frontiers in Genetics* **11**, 189.

Donnelly M.P., Paschou P., Grigorenko E., Gurwitz D., Barta C., Lu R.B., Zhukova O.V., Kim J.J., Siniscalco M., New M., Li H., Kajuna S.L., Manolopoulos V.G., Speed W.C., Pakstis A.J., Kidd J.R. & Kidd K.K. (2012) A global view of the OCA2-HERC2 region and pigmentation. *Human Genetics* **131**, 683–96.

Falker-Gieske C., Blaj I., Preuß S., Bennewitz J., Thaller G. & Tetens J. (2019) GWAS for meat and carcass traits using imputed sequence level genotypes in pooled F2-designs in pigs. *G3: Genes, Genomes, Genetics* **9**, 2823–34.

Ferenčaković M., Hamzić E., Gredler B., Solberg T. R., Klemetsdal G., Curik I. & Sölkner J. (2013a). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics, 130*(4), 286-293.

Ferenčaković M., Sölkner J. & Curik I. (2013b) Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genetics Selection Evolution* **45**, 42.

Fernandes A.C., da Silva V.H., Goes C.P., Moreira G.C.M., Godoy T.F., Ibelli A.M.G., Peixoto, J. De O., Cantão M.E., Ledur M.C., de Rezende F.M. & Coutinho L.L. (2021) Genome-wide detection of CNVs and their association with performance traits in broilers. *BMC Genomics* **22**, 354.

Fontanesi L., Vargiolu M., Scotti E., Latorre R., Faussone Pellegrini M.S., Mazzoni M., Asti M., Chiocchetti R., Romeo G., Clavenzani P. & De Giorgio R. (2014) The *KIT* gene is associated with the *English spotting* coat color locus and congenital megacolon in Checkered Giant rabbits (*Oryctolagus cuniculus*). *PLoS One* **9**, e93750.

Fontanesi L. (2021a) Rabbit genetic resources can provide several animal models to explain at the genetic level the diversity of morphological and physiological relevant traits. *Applied Sciences (Basel)* **11**, 373.

Fontanesi L. (2021b) Genetics and Molecular Genetics of Coat Colour in the European Rabbit. In: Fontanesi L., Ed. *The Genetics and Genomics of the Rabbit*. Wallingford, UK: CAB International; 2021. p. 84–103.

Fontanesi L., Bovo S. & Schiavo G. (2021a) The Genome of the European Rabbit and Genomic Tools. In: Fontanesi L, editor. The Genetics and Genomics of the Rabbit. Wallingford, UK: CAB International; 2021. p. 38–42.

Fontanesi L., Schiavo G., Galimberti G., Calò D.G., Scotti E., Martelli P.L., Buttazzoni L., Casadio R. & Russo V. (2012) A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics* **13**, 583.

Fontanesi L., Utzeri V.J. & Ribani A. (2021b) The Evolution, Domestication and World Distribution of the European Rabbit (*Oryctolagus cuniculus*). In: Fontanesi L., Ed. *The Genetics and Genomics of the Rabbit*. Wallingford, UK: CAB International; p. 1–22.

Gibson J., Morton N.E. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789–95.

Gorssen W., Meyermans R., Buys N. & Janssens S. (2020) SNP genotypes reveal breed substructure, selection signatures and highly inbred regions in Piétrain pigs. *Animal Genetics* **51**, 32–42.

Grilz-Seger G., Druml T., Neuditschko M., Mesarič M., Cotman M. & Brem G. (2019) Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Animal Genetics* **50**, 334–46.

Howard J.T., Pryce J.E., Baes C. & Maltecca C. (2017) Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *Journal of Dairy Science* **100**, 6009–24.

Kardos M., Luikart G. & Allendorf F.W. (2015) Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* **115**, 63–72.

Knief U., Kempenaers B. & Forstmeier W. (2017) Meiotic recombination shapes precision of pedigree- and marker-based estimates of inbreeding. *Heredity* **118**, 239–48.

Kristensen T.N., Pedersen K.S., Vermeulen C.J. & Loeschcke V. (2010) Research on inbreeding in the "omic" era. *Trends in Ecology and Evolution* **25**, 44-52.

Laghouaouta H., Sosa-Madrid B.S., Zubiri-Gaitán A., Hernández P. & Blasco A. (2020) Novel genomic regions associated with intramuscular fatty acid composition in rabbits. *Animals* **10**, 2090.

Leutenegger A.L., Prum B., Génin E., Verny C., Lemainque A., Clerget-Darpoux F. & Thompson E.A. (2003) Estimation of the inbreeding coefficient through use of genomic data. *American Journal of Human Genetics* **73**, 516–23.

Liu C., Wang S., Dong X., Zhao J., Ye X., Gong R. & Ren Z. (2021) Exploring the genomic resources and analysing the genetic diversity and population structure of Chinese indigenous rabbit breeds by RAD-seq. *BMC Genomics* **22**, 573.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L. & Macciotta N.P.P. (2015) Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics* **46**, 110–21.

Mastrangelo S., Sardina M.T., Tolone M., Di Gerlando R., Sutera A.M., Fontanesi L. & Portolano B. (2018) Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal* **12**, 2480–8.

Mastrangelo S., Sardina M.T., Tolone M., Di Gerlando R., Sutera A.M., Fontanesi L. & Portolano B. (2018) Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal* **12**, 2480–8.

Meyermans R., Gorssen W., Buys N. & Janssens S. (2020) How to study runs of homozygosity using plink? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics* **21**, 94.

Muráni, E., Ponsuksili, S., d'Eath, R. B., Turner, S. P., Kurt, E., Evans, G., Thölking L., Klont R., Foury A., Mormède P. & Wimmers K. (2010) Association of HPA axis-related genetic variation with stress reactivity and aggressive behaviour in pigs. *BMC Genetics* **11**, 74.

Paredes-Sánchez F.A., Sifuentes-Rincón A.M., Casas E., Arellano-Vera W., Parra-Bracamonte G.M., Riley D.G., Welsh T.H.Jr. & Randel R.D. (2020) Novel genes involved in the genetic architecture of temperament in Brahman cattle. *PLoS One* **15**, e0237825.

Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R. & Baldi F. (2017) Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics* **48**, 255–71

Purfield D.C., McParland S., Wall E. & Berry D.P. (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One* **12**, e0176780.

Reyer H., Ponsuksili S., Wimmers K. & Murani E. (2014) Association of N-terminal domain polymorphisms of the porcine glucocorticoid receptor with carcass composition and meat quality traits. *Animal Genetics* **45**, 125–9.

Rodríguez-Ramilo S.T., Reverter A., Sánchez J.P., Fernández J., Velasco-Galilea M., González O. & Piles M. (2020) Networks of inbreeding coefficients in a selected population of rabbits. *Journal of Animal Breeding and Genetics* **137**, 599–608.

Sahana G., Kadlecová V., Hornshøj,H., Nielsen B., & Christensen O.F. (2013) A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. *Journal of Animal Science* **91**, 1041–50.

Schiavo G., Bovo S., Muñoz M., Ribani A., Alves E., Araújo J.P., Bozzi R., Čandek-Potokar M., Charneca R., Fernandez A.I., Gallo M., García F., Karolyi D., Kušec G., Martins J.M., Mercat M.J., Núñez Y., Quintanilla R., Radović Č., Razmaite V., Riquet J., Savić R., Usai G., Utzeri V.J., Zimmer C., Ovilo C. & Fontanesi L. (2021) Runs of homozygosity provide a genome landscape picture of inbreeding and genetic history of European autochthonous and commercial pig breeds. *Animal Genetics* **52**, 155–70.

Schiavo G., Bovo S., Bertolini F., Dall'Olio S., Nanni Costa L., Tinarelli S., Gallo M. & Fontanesi L. (2020a) Runs of homozygosity islands in Italian cosmopolitan and autochthonous pig breeds identify selection signatures in the porcine genome. *Livestock Science* **240**, 104219.

Schiavo G., Bovo S., Bertolini F., Tinarelli S., Dall'Olio S., Nanni Costa L., Gallo M. & Fontanesi L. (2020b) Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. *Animal* **14**, 915–20.

Sorbolini S., Bongiorni S., Cellesi M., Gaspa G., Dimauro C., Valentini A. & Macciotta N.P.P. (2017) Genome wide association study on beef production traits in Marchigiana cattle breed. *Journal of Animal Breeding and Genetics* **134**, 43–8.

Sosa-Madrid BS, Hernández P, Blasco A, Haley CS, Fontanesi L, Santacreu MA, Pena RN, Navarro P, Ibáñez-Escriche N. (2020) Genomic regions influencing intramuscular fat in divergently selected rabbit lines. *Animal Genetics* **51**, 58–69.

Sosa-Madrid B.S., Santacreu M.A., Blasco A., Fontanesi L., Pena R.N. & Ibáñez-Escriche N. (2020) A genomewide association study in divergently selected lines in rabbits reveals novel genomic regions associated with litter size traits. *Journal of Animal Breeding and Genetics* 137, 123–38.

Sosa-Madrid B S., Varona L., Blasco A., Hernández P., Casto-Rebollo C. & Ibáñez-Escriche N. (2020c) The effect of divergent selection for intramuscular fat on the domestic rabbit genome. *Animal* **14**, 2225–35.

VanRaden P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–23.

VanRaden P.M., Olson K.M., Null D.J. & Hutchison J.L. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science* **94**, 6153–61.

Wang J. (2016) Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology* **107**, 4–13.

Weng Z., Su H., Saatchi M., Lee J., Thomas M.G., Dunkelberger J.R. & Garrick D.J. (2016) Genome-wide association study of growth and body composition traits in Brangus beef cattle. *Livestock Science* **183**, 4–11.

Whitman B.D. (2004) *Domestic Rabbits & Their Histories: Breeds of the World*. Leathers Publishing. Overland Parks, KS, USA: Leathers Publishing.

Wright S. (1922) Coefficients of Inbreeding and Relationship. *American Naturalist* **56**, 330–8.

Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011) GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82.

Zeder M.A. (2012) The domestication of animals. *Journal of Anthropological Research* **68**, 161–90.

Zeuner F.E. (1963) *A History of Domesticated Animals*. Harper & Row Publishers, Evanston, UK.

Zhou L., Ji J., Peng S., Zhang Z., Fang S., Li L., Zhu Y., Huang L., Chen C. & Ma J. (2016) A GWA study reveals genetic loci for body conformation traits in Chinese Laiwu pigs and its implications for human BMI. *Mammalian Genome* **27**, 610–21.

Table 1. General statistics of runs of homozygosity (ROH) identified in 16 rabbit breeds based on the Method 2 to call ROH. General statistics of ROH detected using the other methods (Methods 1, 3 and 4) are reported in Table S4.

Breed[1]	**nROH (SD)**[2]	**SROH (SD)**[3]	**LROH (SD)**[4]
Belgian Hare (BH)	150.167 (13.160)	735.907 (118.772)	4.887 (0.661)
Burgundy Fawn (BF)	140.500 (7.259)	780.595 (54.878)	5.583 (0.644)
Champagne d'Argent (CdA)	144.579 (9.703)	752.809 (227.838)	5.209 (1.588)
Checkered Giant (CG)	123.853 (26.457)	516.631 (166.303)	4.066 (0.919)
Coloured Dwarf (CD)	124.600 (8.929)	708.714 (199.047)	5.651 (1.369)
Commercial Meat (CM)	120.250 (13.038)	496.797 (92.528)	4.118 (0.617)
Dwarf Lop (DL)	127.800 (12.060)	639.518 (147.462)	5.026 (1.242)
Ermine (ER)	148.100 (12.345)	921.273 (179.834)	6.268 (1.355)
Giant Grey (GG)	131.444 (15.621)	553.177 (151.441)	4.146 (0.767)
Giant White (GW)	115.550 (36.309)	530.500 (196.406)	4.386 (1.128)
Italian Silver (ISI)	105.526 (32.039)	548.743 (254.836)	4.894 (1.543)
Italian Spotted (ISP)	113.946 (16.143)	584.898 (130.293)	5.064 (0.870)
Italian White (ITW)	105.819 (16.797)	427.869 (86.400)	4.001 (0.556)
Rex (RE)	118.389 (25.741)	569.153 (197.209)	4.674 (1.098)
Rhinelander (RH)	133.538 (16.962)	620.754 (166.506)	4.599 (1.005)
Thuringian (TH)	123.667 (44.755)	544.502 (231.073)	4.103 (1.105)

[1] The acronym of the breed used in other figures is reported within brackets.

[2]nROH: the average total number of ROH and the standard deviation (SD) calculated for each breed.

[3] L_{ROH}: the average length of ROH (in Mb) considering all length classes and the standard deviation (SD) calculated for each breed.

[4] S_{ROH}: the average sum of all ROH segments (in Mb) by animals considering all length classes and the standard deviation (SD) calculated for each breed.

Table 2. The mean and standard deviation (in brackets) of genomic inbreeding parameters calculated using different classes of ROH length (F_{ROH1} to F_{ROH16}), identified based on the main method to call ROH (method 2). Table S10 includes results obtained with all four Methods.

Breed	**F_{ROH1}**	**F_{ROH4}**	**F_{ROH8}**	**F_{ROH16}**
Belgian Hare	0.377 (0.061)	0.270 (0.065)	0.150 (0.052)	0.042 (0.029)
Burgundy Fawn	0.400 (0.028)	0.306 (0.040)	0.196 (0.055)	0.072 (0.037)
Champagne d'Argent	0.386 (0.117)	0.282 (0.0132)	0.177 (0.129)	0.080 (0.090)
Checkered Giant	0.265 (0.085)	0.164 (0.077)	0.081 (0.061)	0.027 (0.037)
Coloured Dwarf	0.363 (0.102)	0.280 (0.109)	0.189 (0.103)	0.092 (0.074)
Commercial Meat	0.255 (0.047)	0.159 (0.405)	0.044 (0.032)	0.019 (0.025)
Dwarf Lop	0.328 (0.076)	0.239 (0.087)	0.143 (0.087)	0.060 (0.055)
Ermine	0.472 (0.092)	0.381 (0.104)	0.263 (0.113)	0.122 (0.079)
Giant Grey	0.284 (0.078)	0.178 (0.077)	0.089 (0.062)	0.032 (0.034)
Giant White	0.272 (0.101)	0.179 (0.081)	0.104 (0.069)	0.049 (0.47)
Italian Silver	0.281 (0.131)	0.205 (0.120)	0.139 (0.100)	0.064 (0.062)
Italian Spotted	0.300 (0.067)	0.217 (0.064)	0.136 (0.050)	0.054 (0.033)
Italian White	0.219 (0.044)	0.133 (0.037)	0.065 (0.030)	0.022 (0.019)
Rex	0.292 (0.101)	0.206 (0.097)	0.122 (0.075)	0.040 (0.034)
Rhinelander	0.318 (0.085)	0.217 (0.088)	0.116 (0.073)	0.038 (0.048)
Thuringian	0.279 (0.118)	0.182 (0.096)	0.097 (0.061)	0.030 (0.024)

Table 3. Runs of homozygosity islands identified in Checkered Giant, Commercial Meat, Italian Spotted and Italian White breeds.

OCU[1]	Position[2]	Breed	No. of SNPs[3]	No. of genes[4]	Candidate genes (reference)[5]
1	159375514_160629747	Italian White	64	5	*NELL1* (Falker-Gieske *et al.* 2019)
2	127462270_128281278	Checkered Giant	43	3	*VRK2, FANCL* (Paredes-Sánchez *et al.* 2020)
2	161490031_163738094	Italian Spotted	54	5	*TRIB2* (Brunes *et al.* 2021; Fernandes *et al.* 2021)
3	21908055_22417201	Checkered Giant	23	6	-
3	24713065_25428869	Checkered Giant	45	4	*NR3C1* (Muráni *et al.* 2010; Reyer *et al.* 2014)
3	78265347_81591625	Italian Spotted	42	10	*ARMC1* (Zhou *et al.* 2016)
3	140599855_142305209	Checkered Giant, Commercial meat, Italian Spotted, Italian White	77, 75, 70, 72	8	*TRIB1* (Brunes *et al.* 2021)
3	149885379_151645140	Checkered Giant	60	4	*ZFAT* (Grilz-Sege *et al.* 2019)-
4	26298302_28844406	Italian Spotted	71	9	-
4	75359078_76496009	Commercial Meat	60	4	*ELK3* (de Lima *et al.* 2020)
7	47708504_49611011	Italian White	26	10	*IFRD1* (Sorbolini *et al.* 2017)
7	54842373_57319254	Checkered Giant, Commercial Meat	77, 152	3	-
7	140627212_142319237	Checkered Giant, Commercial Meat, Italian Spotted	19, 55, 32	14	*AOX1* (Casto-Redollo *et al.* 2020)
10	5642931_6250075	Commercial	14	3	*ITGB8* (Casto-Redollo *et al.* 2020)
12	8595729_11116965	Italian Spotted, Italian White	48, 48	143	*GPX5* (Barranco *et al.* 2016)
12	78307373_84406018	Italian White, Commercial Meat	60, 15	12	*NDUFAF4* (An *et al.* 2018)
13	74396836_76411306756933348	Italian Spotted, Italian White	39, 54	18	*PKN2* (Fontanesi *et al.* 2012)

15	12623264_15388967	Italian White, Commercial Meat, Italian Spotted	59, 56, 52	12	*DCLK2* (Sahana *et al*. 2013)
15	63477042_65101796	Italian Spotted	58	4	*WDFY*3 (Chang *et al*. 2018)
15	84153011_85988883	Commercial Meat, Italian White	7, 57	3	*TECRL* (Weng *et al*. 2016)
16	14715459_16455125	Checkered Giant	53	5	-
17	75881549_77977324	Checkered Giant	76	20	*OCA2* (Donnelly *et al*. 2012)

[1] *Orycolagus cuniculus* chromosome.

[2] Starting and ending nucleotide position of the ROH region. Partially overlapped ROH in different breeds have been combined in a single ROH region.

[3] Number of SNPs in the ROH identified in the corresponding breeds.

[4] Number of genes annotated in the OryCun2.0 genome version included in the ROH region.

[5] Candidate genes, potentially affecting production traits (like growth, carcass and meat traits, feed efficiency, reproduction traits, body size, behaviour and related traits) and other exterior traits, as identified from a literature survey in other livestock species (the cited references report the effects of the polymorphisms in the indicated genes). *NELL1*: carcass traits; *VRK2, FANCL*: behaviour; *TRIB2*, feed efficiency, growth rate; *NR3C1*: carcass composition, meat quality traits and *stress response; ARMC1: fat deposition; TRIB1: feed efficiency; ZFAT: body size;* ELK3: feed efficiency; *IFRD1*: growth rate; *AOX1* and *ITGB8*: litter size (in rabbit); *GPX5*: male reproduction trait; *NDUFAF4*: organ size; *PKN2*: fat deposition; *DCLK2*: feed efficiency; *WDFY3*: carcass traits; *TECRL*: fat deposition; *OCA2*: pigmentation.

Figure 1. Correlation plots between nROH (y axis) and S_{ROH} (x axis) for the 16 rabbit breeds including all animals. Acronyms of the breeds and are defined in Table 1. Pearson correlation coefficient between nROH and S_{ROH} is reported beside the acronym of each breed.

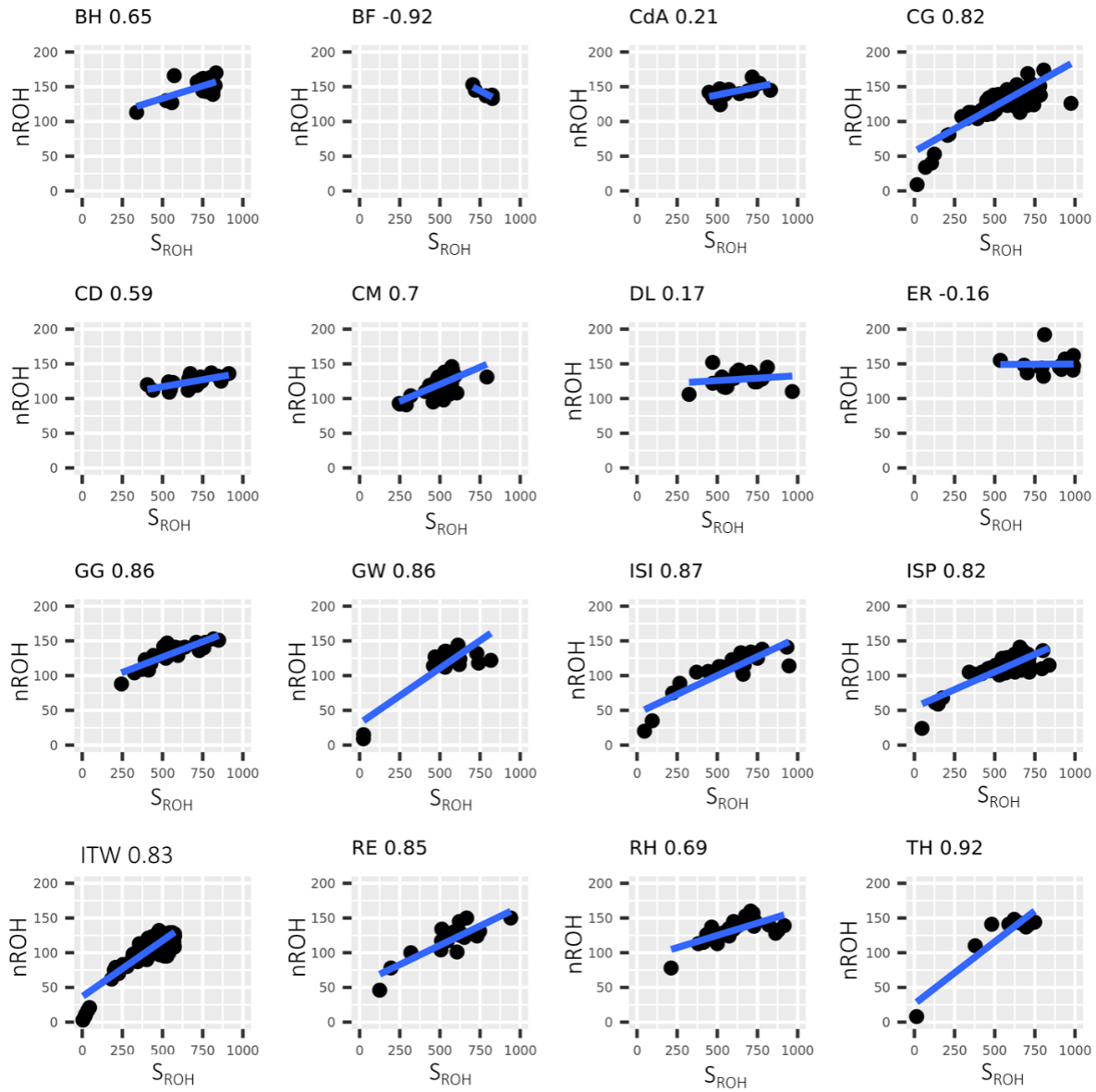


Figure 2. Proportion of runs of homozygosity in 16 rabbit breeds of different ROH categories were defined according to physical size: 1–2, 2–4, 4–8, 8–16 and >16 Mb, identified as ROH1–2, ROH2–4, ROH4–8, ROH8–16 and ROH>16, respectively.

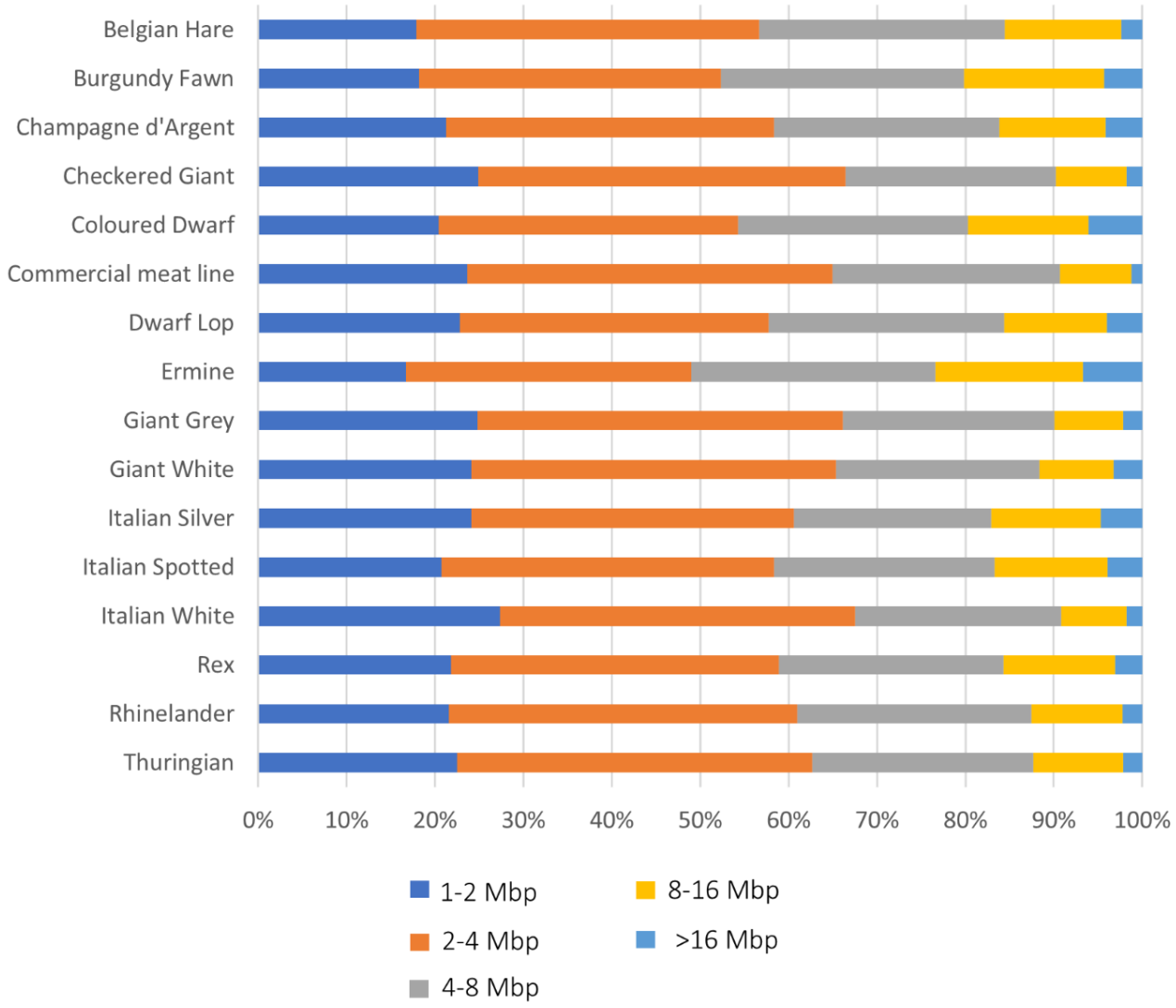
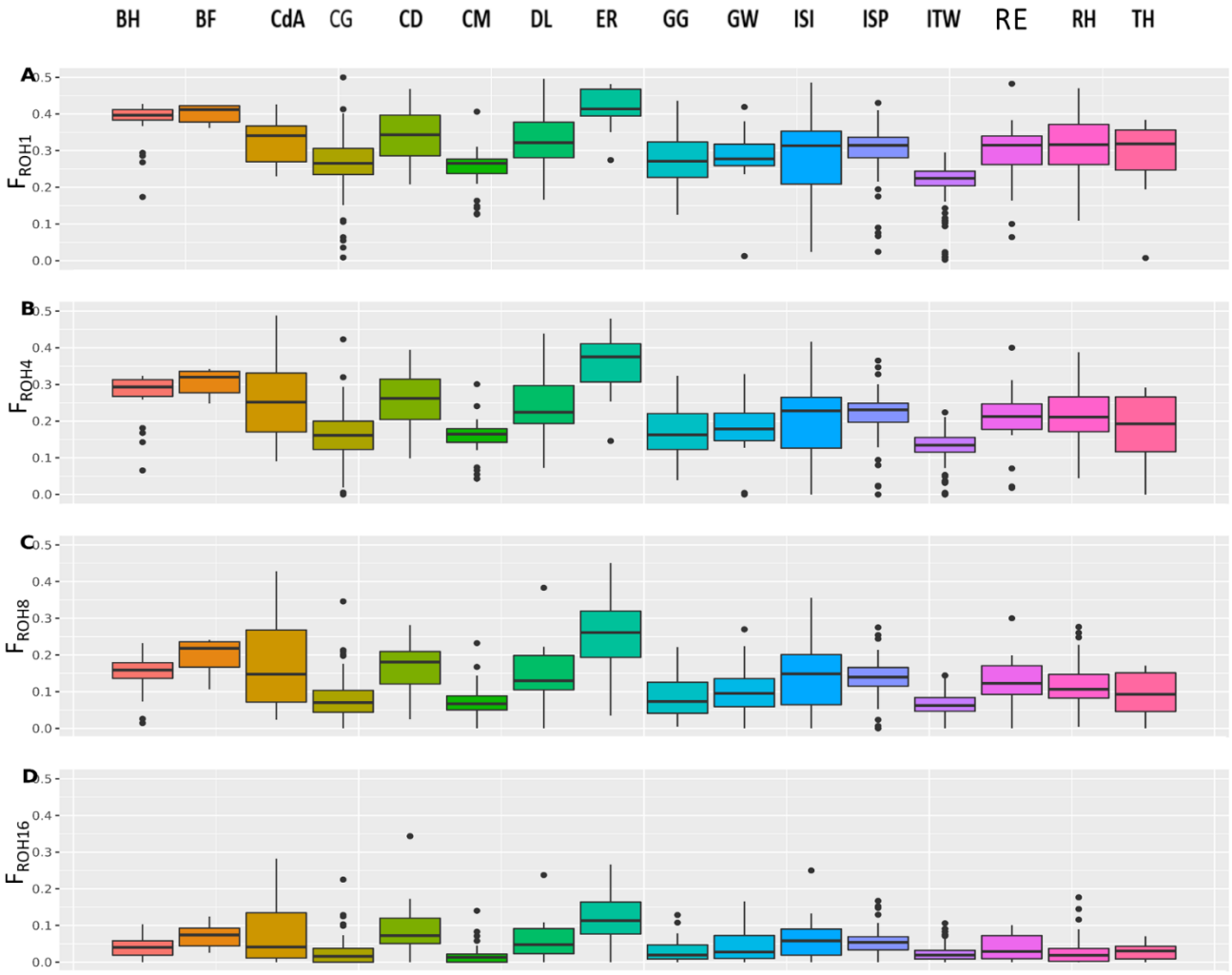
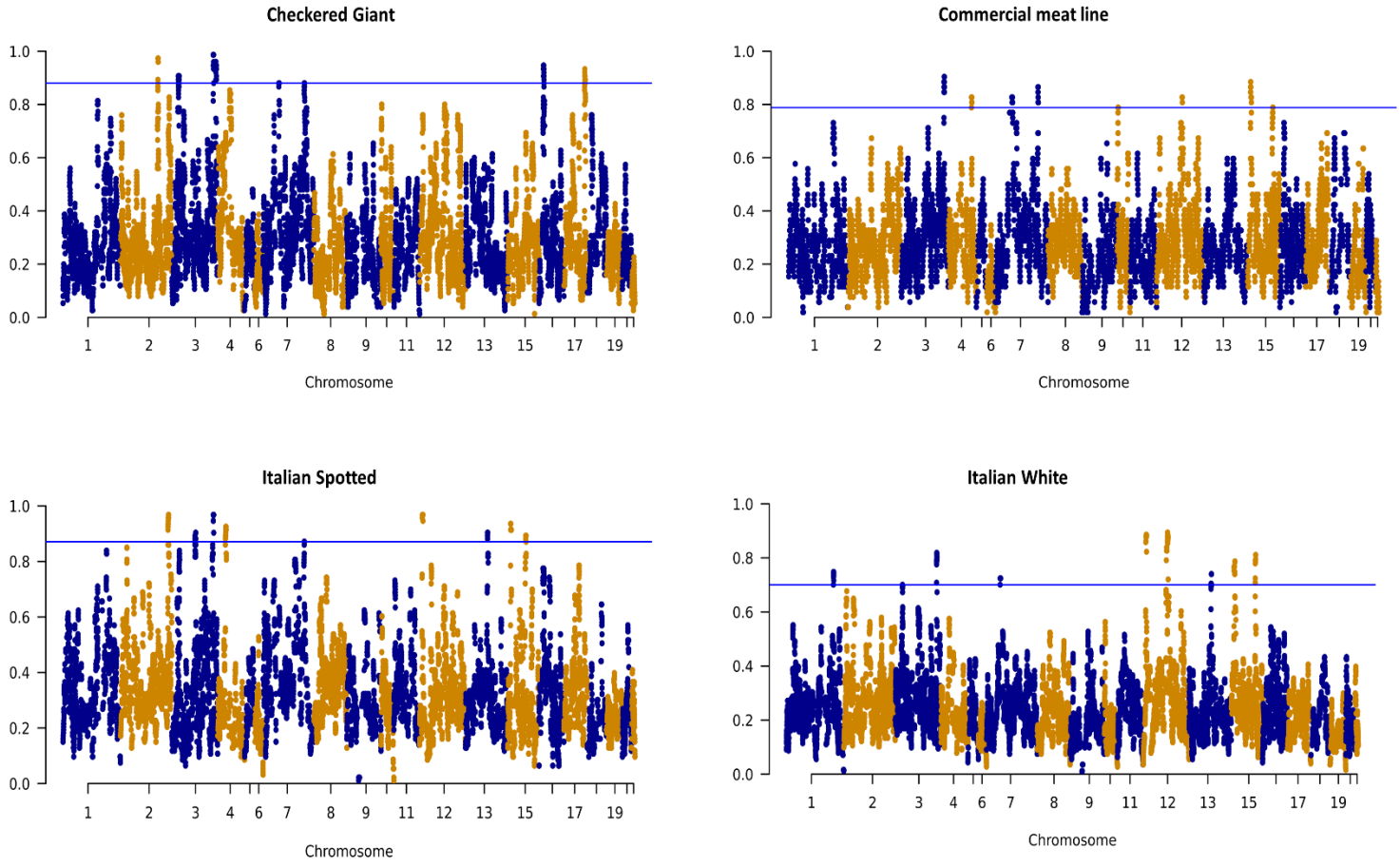


Figure 3. Boxplots of the F_{ROH} distribution in the 16 rabbit breeds: a) F_{ROH1}; b) F_{ROH4}; c) F_{ROH8};



d) F_{ROH16}. Acronyms of the breeds are explained in Table 1.

Figure 4. Manhattan plots showing ROH islands in Checkered Giant, Commercial Meat, Italian Spotted and Italian White breeds. The blue line indicates the frequency corresponding to the top



1% of most frequent SNPs in the population.

Chapter 5 : Population genomic structures and signatures of selection define the genetic uniqueness of several fancy and meat rabbit breeds

Mohamad Ballan[1], Samuele Bovo[1], Francesca Bertolini[1], Giuseppina Schiavo[1], Michele Schiavitto[2], Riccardo Negrini[3], Luca Fontanesi[1,*]

[1]Division of Animal Sciences, Department of Agricultural and Food Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy

[2]Associazione Nazionale Coniglicoltori Italiani (ANCI), Contrada Giancola snc, 71030 Volturara Appula (FG), Italy

[3]Associazione Italiana Allevatori, Via G. Tomassetti 9, 00161 Roma, Italy

* Correspondence to LF

E-mail: luca.fontanesi@unibo.it

Published in Journal of Animal Breeding and Genetics. 12 July 2023.

Doi: https://doi.org/10.1111/jbg.12818

ABSTRACT

Following the recent domestication process of the European rabbit (*Oryctolagus cuniculus*), many different breeds and lines, distinguished primarily by exterior traits such as coat colour, fur structure, and body size and shape, have been constituted. In this study, we genotyped, with a high

density single nucleotide polymorphism panel, a total of 645 rabbits from 10 fancy breeds (Belgian Hare, Champagne d'Argent, Checkered Giant, Coloured Dwarf, Dwarf Lop, Ermine, Giant Grey, Giant White, Rex, and Rhinelander) and three meat breeds (Italian White, Italian Spotted, and Italian Silver). We investigated their population genomic structures using the ADMIXTURE approach and analysed signatures of selection using two haplotype-based approaches (iHS and XP-EHH). To obtain a comprehensive landscape picture of signatures of selection in the same rabbit breeds, we also compared the results of this study with the signatures of selection identified in previous studies detected with other methods (Ballan et al., 2022a, 2022b).. ADMIXTURE analysis indicated that breeds with similar phenotypic traits (e.g., coat colour and body size) shared common ancestries. Then, combining the results obtained in this study and those derived with other methods previously reported, we identified a total of 5079 independent genomic regions with some signatures of selection, covering about 1777 Mb of the rabbit genome. These regions consistently encompassed many genes involved in pigmentation processes (*ASIP*, *EDNRA*, *EDNRB*, *KIT*, *KITLG*, *MITF*, *OCA2*, *TYR,* and *TYRP1*), coat structure (*LIPH*) and body size, including two main genes (*LCORL* and *HMGA2*) among many others. This study revealed novel genomic regions under signatures of selection and further demonstrated that population structures and signatures of selection, left into the genome of these rabbit breeds, may contribute to understanding the genetic events that led to their constitution and the complex genetic mechanisms determining the broad phenotypic variability present in these untapped rabbit genetic resources.

KEYWORDS: Admixture; Animal genetic resources; Domestication; *Oryctolagus cuniculus*; Selection sweep; SNP.

1 INTRODUCTION

The domestication of the European rabbit (*Oryctolagus cuniculus*), known simply as rabbit, has been a recent process, compared to the domestication of other livestock species. It began just over the High Middle Ages in French monasteries and castles using wild populations of the *O. c. cuniculus* subspecies spread in this region after its post-glacial expansion from the Iberian Peninsula (Zeuner, 1963; Callou, 2003; Zeder, 2012; Fontanesi et al., 2021). Then, the domestication process continued in the XV-XVII centuries, with the subsequent dispersion and transport of semi-domesticated/domesticated rabbit populations in the North of Europe and, subsequently, in other European regions (Fontanesi et al., 2021). A genetic bottleneck occurred during this phase due to the use of wild subpopulations and to the limited recurred introgression from the wild forms into the domesticated lineage (Carneiro et al., 2014; Alves et al., 2015). The first waves of this process produced a soft allele frequency modification at many loci distinguishing wild and domestic rabbit populations. This allelic shift mainly occurred in regulatory regions of genes that contributed to the behavioral and reproductive adaptation of the domesticated rabbit stocks to the human environment and needs (Carneiro et al., 2014). Then, the human driven selection contributed to domestication with the constitution of the first rabbit breeds (Boucher et al., 2021; Fontanesi, 2021a). The subsequent result was that a variety of peculiar phenotypic differences were fixed in many breeds which were valued by fancy breeders who continued the selection activities and then created additional breeds, lines, and strains by introgressing specific genetic factors to obtain novel genetic combinations (Boucher et al., 2021; Fontanesi, 2021a). Most of the fancy breeds were named according to their characterizing exterior traits (mainly derived by their coat colours and patterns, body size, ear length, and pelage features), which became their specific breed standards defined by breeders' associations or societies,

constituted in several countries (Whitman, 2004; Boucher et al., 2021). Nowadays, some of these fancy breeds are recognized by all or almost all national breeders' organizations, and a few other breeds, mainly defined by coat colour variants, are recognized only in one or few countries. Specialized meat rabbit lines were more recently developed through the selection activities of breeding companies that mainly valued carcass, growth, and reproduction traits.

The selection histories and the genetic events that led to the constitution of many different fancy breeds and meat lines have defined population genetic structures (Alves et al., 2015) and left signatures of selection in their genomes that made possible to identify the genetic factors affecting some of the most relevant phenotypes that characterize several rabbit genetic resources (Fontanesi, 2021a). Using a combination of the candidate gene approach and linkage analysis, we and others identified the causative mutations or the associated markers that explain the allele series at several coat colour loci, including the *Agouti, Albino*, *Brown, Dilute*, *English Spotting,* and *Extension* loci (Aigner et al., 2000; Lehner et al., 2013; Fontanesi et al. 2006, 2010a, 2010b, 2014a, 2014b; Utzeri et al., 2014, 2021; Demars et al., 2018; Letko et al., 2020), and at other loci affecting, for example, the rex hair coat phenotype (Diribarne et al., 2011), a type of dwarfism (Carneiro et al., 2017), and a saltatorial behavior (Carneiro et al., 2021). High density single nucleotide polymorphisms (SNPs) covering the whole rabbit genome have been used to identify several other genomic regions under selection in fancy and meat rabbit lines using some classical approaches based on F_{ST} and runs of homozygosity (ROH) islands (Ballan et al., 2022a, 2022b).

Several other methods and approaches useful to identify signatures of selection have been developed based on high density single markers or haplotypes. Many applications in human and animal populations have already demonstrated their complementarities and usefulness to further confirm some previously identified signatures of selection (as proof of concepts) and in extracting

additional signatures of selection that could also characterize the populations under investigation (e.g.: Sabeti et al., 2006; Utsunomiya et al., 2013; Qanbari and Simianer, 2014; González-Rodríguez et al., 2016; Saravanan et al., 2020). For example, two frequently applied haplotype-based tests, considered to be complementary to capture signatures of selection (Gautier and Naves, 2011), are the integrated haplotype score (iHS; Voight et al., 2006), which derives from the extended haplotype homozygosity (EHH), and the cross-population extended haplotype homozygosity (XP-EHH; Sabeti et al., 2007), which is based on both EHH and iHS principles and that further extend the possibility to detect haplotype derived information at the genome level.

In this study, we first analysed their population genomic structures using high density SNP datasets obtained from 13 rabbit breeds (10 fancy breeds and three meat breeds or lines). Then we applied two haplotype-based approaches (iHS and XP-EHH) to identify signatures of selection in their genomes with a combination of within breed and across breeds analyses. The results were then compared with the signatures of selection that were previously obtained with other methods to compile a more comprehensive pattern of selection sweeps that distinguishes rabbit breeds and that might be useful to reconstruct the genetic history of these animal genetic resources and identify the genetic mechanisms that can explain their broad phenotypic diversity.

2 MATERIALS AND METHODS

2.1 Ethic statement

Animal samples used in this study were collected following the recommendation of directive 2010/632.1.

2.2 Animals

Hair roots or buccal swabs were sampled from 645 rabbits from 10 fancy and three meat rabbit breeds. All fancy breeds (Belgian Hare, n. 24; Champagne d'Argent, n. 19; Checkered Giant, n. 79; Coloured Dwarf, n. 20; Dwarf Lop, n. 20; Ermine, n. 20; Giant Grey, n. 27; Giant White, n. 20; Rex, n. 19; and Rhinelander, n. 28) and meat breeds (Italian White, n. 256; Italian Spotted, n. 93; Italian Silver, n. 20) were from the national Herd Book maintained by the Italian Rabbit Breeders Association (ANCI). All rabbits included in this study were selected to avoid highly related animals (no full- or half-sibs) and had their respective standard breed characteristics. The description of the breeds is reported in Ballan et al. (2022a) and summarized in Table S1.

To capture signature of selection common to more than one breed with similar characteristics, breeds were also grouped according to a few common features (Table S1 and Table S2) based on coat colour (completely white *Albino* breeds: Giant White and Italian White; silver breeds: Champagne d'Argent and Italian Silver; spotted breeds: Checkered Giant and Rhinelander), body size (giant breeds: Checkered Giant, Giant Grey, and Giant White; dwarf/small breeds: Coloured Dwarf, Dwarf Lop, and Ermine) and the use of the breed (meat breeds: Italian White, Italian Spotted, and Italian Silver; fancy breeds: all other breeds).

2.3 Genotyping, data filtering and phasing

DNA was extracted using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). Animals were then genotyped using the Affymetrix Axiom OrcunSNP Array (Affymetrix Inc., Santa Clara, CA, USA) which can analyse 199692 DNA markers. The genotyping data quality check was made with the Axiom Analysis Suite and PLINK v.1.9 software (Chang et al., 2015). SNPs with a call rate < 0.90, unmapped or on sex chromosomes were removed and after filtering, a total number of 139922 SNPs from 645 rabbits were retained for further

analyses. Then, the SNP dataset of all investigated rabbits was phased using SHAPEIT2, with the default parameters (Delaneau et al., 2012).

2.4 Population structure analyses

To investigate the relationships between the analyzed breeds, multidimensional scaling (MDS) analysis was performed after obtaining the matrix of genome-wide identity by state (IBS) pairwise distances using PLINK 1.9 (Chang et al., 2015). ADMIXTURE 3.1 software (Alexander et al., 2009) evaluated population stratification. Following (Bertolini et al., 2022), the pruning of SNPs in high LD was carried out using PLINK 1.9 (Chang et al., 2015) considering –indep-pairwise command with the default parameters (windows size of 1000 kb and r^2 threshold of 0.2). A total of 5858 SNPs were retained (an average of 163 SNPs for each chromosome). Analysis was performed on the pruned dataset considering the number of sub-populations (K) that ranged from 1 to 17 and calculating the cross-validation error (CV) for each K.

2.5 Haplotype-based methods to detect signatures of selection

The rehh R package V 2.04 (Gautier et al., 2017) was used on the phased SNPs to obtain the integrated haplotype scores (iHS; Voight et al., 2006) for each breed and across-population and the extended haplotype homozygosity (XP-EHH; Sabeti et al., 2007) values for breed comparison and groups of breed comparisons. iHS values were calculated with the unpolarized option, available for this method when it is impossible to define the ancestral status of the alleles at a polymorphic site. Here, absolute values |iHS| were chosen and considered independent from the ancestral allele (Kemper et al., 2014; Bertolini et al., 2020). Signatures of selection were identified

based on these absolute values for SNPs that passed the threshold of the top 98.0th percentile of the empirical distribution with at least three consecutive SNPs within a 350 kb window.

The XP-EHH analysis is based on a single breed approach, considered for each breed N-1 pairwise XP-EHH comparisons (where N is the number of breeds included in the study; N=13). The final XP-EHH score for each SNP was then calculated for each breed, considered the reference population, by computing the mean of the XP-EHH values obtained from these multiple comparisons. XP-EHH analyses for groups of breeds were carried out considering as one reference population all rabbits belonging to the grouped breeds against another population, including all rabbits of the other breeds (Table S2). Therefore, only negative XP-EHH mean values were considered (Maiorano et al., 2018). To detect a signature of selection, all SNPs within the top 95.0th percentile of the distribution of their log P-value were retained. Then, regions with at least three consecutive SNPs within a 350 kb window with negative XP-EHH values were considered to identify signatures of selection with this approach.

2.6 Compilation of signatures of selection in the rabbit genome

Signatures of selection identified with the iHS and XP-EHH analyses derived from this study were combined with signatures of selection obtained in previous studies for the same rabbit breeds genotyped with the same SNP array. These additional signatures of selection were obtained using other approaches: window based F_{ST}, a principal component analysis-based method obtained with *PCAdapt* (Luu et al., 2017), and ROH islands (Ballan et al., 2022a, 2022b). Genome coordinates of all these regions were obtained using OryCun2.0 chromosomes. Annotated protein-coding genes within the reported genome windows or that were ± 500 kb apart from them were retrieved with the bedtools v.2.17 (https://github.com/arq5x/bedtools2/) from the OryCun2.0 NCBI's GFF

file (NCBI Oryctolagus cuniculus Annotation Release 102). Gene enrichment analyses were performed with Enrichr (Chen et al., 2013), using the following databases: GWAS catalog (https://www.ebi.ac.uk/gwas/), KEGG Human database (KEGG, http://www.kegg.jp/), MGI mammalian phenotype level (https://www.informatics.jax.org/vocab/mp_ontology), and the biological process branch of gene ontology (GO:BP; http://geneontology.org/). As input, Enrichr took the whole set of genes (n = 1493) mapped within the genome regions identified by more than one method. We considered statistically enriched terms presenting: (i) at least two genes of the input set related to (at least) two different genome regions and (ii) an adjusted *P*-value < 0.05.

3 RESULTS

3.1 Population structure of the investigated breeds

The MDS plots, obtained with the genotyped SNPs, including the rabbits of the 13 breeds, are reported in Figure 1. Rabbits belonging to the same breeds were usually clustered, with few exceptions when breeds were not separated. The Italian White breed was clustered apart from all other breeds. Breeds grouped according to the same type of body size, i.e. the three giant breeds (Checkered Giant, Giant Grey, and Giant White) and the three dwarf/small breeds (Coloured Dwarf, Dwarf Lop, and Ermine), and the two breeds with the silver coat colour (Champagne d'Argent, and Italian Silver), were clustered closely together.

Figure 2 shows the results of the ADMIXTURE analysis. Within the 13 subpopulations, the lowest value was obtained at K 17, while the CV error reached a first plateau at K 11. At K 11, the three meat breeds (Italian White, Italian Spotted, and Italian Silver) shared some ancestries which were maintained, only in part, at K 13 (mainly between Italian Spotted and Italian Silver) and K 17 (mainly between Italian White and Italian Spotted). The two silver breeds (Italian Silver

and Champagne d'Argent) shared large common ancestries at K 11 and K 13 and a lower level at K 17. An admixed genetic background was observed between Giant White and Giant Grey at all three mentioned Ks, with a much lower level of admixture in Checkered Giant. The three small/dwarf breeds (Colored Dwarf, Dwarf Lop, and Ermine) shared some common ancestries at K 11. At K 13, common ancestries were high between Coloured Dwarf and Dwarf Lop and between Ermine and Coloured Dwarf; the latest was also maintained at K 17. Some common ancestries also emerged between the two dwarf breeds and the Rex breed at K 11 which was then reduced at K 13. No common ancestries emerged for Rhinelander and Belgian Hare at all three Ks.

A very low level of introgression at the same three Ks was shown by these two breeds, i.e. Belgian Hare and Rhinelander. At K 13, in addition to the previously mentioned breeds, Champagne d'Argent, Ermine and Dwarf Lop showed the lowest levels of introgression, followed by Giant Grey and Giant White (that were genetically very similar at this level) and, subsequently, by Rex. All three meat breeds showed two main groups of different ancestries at K 11 and K 13, which also remained or increased at three groups (for Italian White) at K 17. The Coloured Dwarf breed consisted of heterogeneous ancestries, with two main groups at K 13 and K 17.

3.2. Signatures of selection identified with the integrated haplotype score |iHS| analyses

Figure S1a shows the Manhattan plots reporting the |iHS| values obtained for all single breeds. The Manhattan plots obtained for the breeds grouped according to their coat colour, body size, and use (i.e. meat breeds) are included in Figure 3a. The list of the top 20 genomic windows for |iHS| values identified with the analyses based on single-breeds (8 out of 20) or groups of breeds (13 out of 20) is reported in Table 1 (one top value was identified in both analyses). The

complete overview of the |iHS| values across single breeds and groups of breeds is reported in Table S3 and Table S4, respectively.

The top 20 |iHS| windows were located on OCU1, OCU2, OCU4, OCU7, OCU9, OCU12, OCU15, and OCU16. Many genes involved in growth processes and body morphology are included in these windows. For example, the top |iHS| value, identified in the Italian White breed and in the meat breeds group, highlighted window on OCU2 that encompasses the *ligand dependent nuclear receptor corepressor like* (*LCORL*) gene, which is well known to affect stature and body size in mammals (e.g. Pryce et al., 2011). Another top ranked genomic region of this list, identified on OCU4 in the Checkered Giant breed, contains the *high mobility group AT-hook 2* (*HMGA2*) gene, which is involved in several basic biological processes. Including mesenchymal differentiation, adipogenesis and post-natal myogenesis. The genomic region encompassing this gene was also evidenced in the Dwarf Lop breed (Figure S1a) and in the analyses involving two breeds groups: giant breeds and dwarf breeds (Figure 3a). Several other relevant genes in the detected |iHS| play roles in growth-related processes and meat production traits (Table S3 and Table S4), some of which are also annotated in Figures 3a and S1a. We can mention a few of them, among several others: *discoidin domain receptor tyrosine kinase* 2 (*DDR2*) that plays a role in signal transduction pathways involved in cell adhesion, proliferation, and extracellular matrix remodeling; *gonadotropin releasing hormone receptor* (*GNRHR*), that is involved in abnormality of body height; *SET domain containing 7, histone lysine methyltransferase* (*SETD7*), that plays a central role in the transcriptional activation of several genes; *suppressor of cytokine signaling 2* (*SOCS2*), that is involved in mechanisms affecting body height.

Other genomic windows containing genes affecting pigmentation were identified in the analyses. Including groups of breeds (*endothelin receptor type A* or *EDNRA*; *endothelin receptor*

type B or *EDNRB*; and *KIT proto-oncogene receptor tyrosine kinase* or *KIT;* Figure 3a) and in the single breed analyses (*agouti signaling protein* or *ASIP* in Belgian Hare and Giant Grey; *EDNRA* in Checkered Giant; *EDNRB* in Champagne d'Argent, Coloured Dwarf, Giant White, Italian Spotted and Italian White; *KIT* in Champagne d'Argent, Ermine, Italian Spotted, Italian White; *OCA2 melanosomal transmembrane protein* or *OCA2* in Checkered Giant; *paired box 2* or *PAX2* in Giant White; *tyrosinase-related protein 1* or *TYRP1* in Dwarf Lop; Figure S1a).

3.3. Signatures of selection identified with the *XP-EHH* analyses

Results of the single breed XP-EHH pairwise comparisons are presented in Table S5 and are shown in the Manhattan plots of Figure S1b. Figure 3b and Table S6 report analyses' results carried out using the defined groups of breeds. Based on the lists of obtained windows, the results of the XP-EHH analyses were in part overlapping and part complementary with the |iHS| results. For example, the *HMGA2* gene region, already reported in the |iHS| results, emerged again in the Checkered Giant and spotted breed group. Other regions, which included many genes involved in growth, body size and development (according to the information available in humans and other species; e.g. *collagen type XI alpha 2 chain, COL11A2*; *non-SMC condensin II complex subunit G2, NCAPG2*; *semaphorin 4D, SEMA4D*; *SMAD family member 1, SMAD1*; *transforming growth factor beta receptor 2, TGFBR2*; *zinc finger and AT-hook domain containing, ZFAT*), emerged from the within breed comparisons and the analyses based on groups of breeds with this method but not with the iHS method (Figure 3b and Figure S1b).

A few regions containing genes affecting pigmentation, identified in the |iHS| analyses, also emerged with the XP-EHH analyses: *KIT* in spotted breeds, Checkered Giant and Ermine; *OCA2* in Checkered Giant; *tyrosinase* (*TYR*) in Italian White (and also in Italian Spotted and the Albino

breed comparison, not emerged with the previous method); *EDNRB* in Belgian Hare (a breed that however did not show this region with the previous method). Other regions containing relevant genes involved in pigmentation emerged only with the XPP-EHH method: KIT ligand (*KITLG*) in albino breeds and Italian White; *melanocyte inducing transcription factor* (*MITF*) in Champagne d'Argent.

Another expected signature of selection was identified in the Rex breed in the genomic region of OCU14 that harbors the *lipase H* (*LIPH*) gene, which is responsible for the Rex^1 locus affecting coat structure (Diribarne et al., 2011).

3.4. Compilation of signatures of selection in the genome of the investigated rabbit breeds

To obtain a complete picture of the signatures of selection identified in the investigated rabbit breeds, we combined results obtained in this study using the iHS and the XP-EHH approaches with the results we previously obtained with other methods, including different F_{ST} pairwise analyses, the application of the principal component analysis methodology of *PCAdapt* (Luu et al., 2017) and the identification of runs of homozygosity (ROH) islands (Ballan et al., 2022a, 2022b). Figure 4 shows an overview of the genomic regions covered by signatures of selection identified by three, four or five different methods. The complete list of all genomic regions identified with one or more methods in different breeds is reported in Table S7. Considering all methods and all breeds together, a total of 5079 independent regions that might have been under selection were identified, covering a total of 1777 Mb. The most consistent results obtained with three, four or five methods, covered a total of 516, 118 and 22 Mb respectively.

In the single breed analyses, 12 genomic regions were identified by at least four different methods (Table 2). Again, a few of these regions consistently highlighted some obvious candidate

genes affecting breed-specific traits. For example, four different methods made possible to identify in the Italian White breed (an albino breed) the genomic region on OCU1 containing the *TYR* gene whose allele series is responsible of the *Albino* locus (Fontanesi, 2021b). Several other regions might harbor candidate genes explaining other breed-specific phenotypes or domestication derived traits. The region on OCU3, identified in Giant White, has been associated with the number of teats in rabbits (Bovo et al., 2021). One region of OCU4 identified in Dwarf Lop includes the *kinesin family member 16B* (*KIF16B*) gene involved in early embryonic development and body size (Ueno et al., 2011; Yengo et al., 2022). A region on OCU9 identified in Checkered Giant includes the *protein tyrosine phosphatase non-receptor type 2* (*PTPN2*) gene, which also emerged as one of the most relevant peaks in the *PCAdapt* analysis based on all meat rabbit breeds (Ballan et al., 2022a). Another consistent region that emerged in Checkered Giant was on OCU17 and included a coat colour gene, the *OCA2* gene.

The enrichment analysis's results, based on the genes annotated in the chromosome regions where at least two methods detected signatures of selection, are reported in Table S8. Just two GO terms, identified in the Checkered Giant breed (the axial length) and the Giant Grey breed (eye color), were significantly enriched.

4. DISCUSSION

The more recent outcome of the domestication process in the rabbit has been the constitution of many different breeds that can be distinguished mainly through their exterior traits, like coat colour and structure, body size, and shape (Boucher et al., 2021; Fontanesi, 2021a). These genetic resources, mainly established and maintained by fancy and commercial breeders, may contain the footprints that testify their genetic history and origin in their genome. Several genetic events and

breeding practices, including artificial directional selection, bottleneck, genetic drift, and introgression, have shaped the genomic architecture and population structures of these breeds. Identifying the signatures of selection left into the genome by the combined action of these events may be useful to understand the genetic mechanisms that determine the peculiar phenotypic characteristics of many of these largely untapped genetic resources.

We already reported some population genomic information and the identification of signatures of selection in several rabbit breeds using a few approaches (F_{ST} methods, an adapted PCA based approach and,

ROH islands) that made possible to identify or confirm genetic hints associated with some relevant phenotypes in the domestic rabbit (Ballan et al., 2022a, 2022b). Here, we reported additional layers of information obtained by further exploiting high-density genotyping data in the same breeds. Our population structure investigation that we carried out with SNP genotyping data enlarged the population structure analysis based on microsatellite information carried out in 16 rabbit breeds by Alves et al. (2015), who involved just a few breeds in common with our study (Belgian Hare, Champagne d'Argent or Champagne Silver, and Rex) or that might be considered similar or with a close origin respect to our investigated breeds (Netherland Dwarf similar to Coloured Dwarf; Flemish Giant and Hungarian Giant similar to Giant Grey). It is also worth mentioning that the samples animals in the study of Alves et al. (2015) were from France and other countries, but not Italy.

Admixture analysis that we carried out in our 13 investigated breeds clearly showed that the breeds with similar phenotypic characteristics (body size and coat colour) and, to some extent, the use for meat production, also shared common genetic features that can affect both quantitative traits and monogenic or oligogenic traits, which, in turn, differentiate their populations. This aspect

did not emerge from the study of Alves et al. (2015), where breeds could not be grouped based on some common features and the applied methodology did not report any results on macro-groups of breeds.

In our study, two giant breeds (Giant Grey and Giant White) largely shared the same ancestral genetic features that could only be distinguished at the subpopulation level (K = 17). The third giant breed, Checkered Giant, was quite divergent from the other two. The main characteristic of this breed is its unique spotted coat colour derived from the heterozygous genotype at the *English spotted* locus (Fontanesi et al., 2014b; Fontanesi, 2021b). Therefore, only planned crosses that follow the classical Mendelian inheritance can produce the required phenotype in the rabbits which, in turn, can respect this breed standard. These needed practices may also contribute, on hand, separatating this giant breed from the two other breeds and, on the other hand, to introduce some within genetic heterogeneity, as evidenced at K= 13 and K = 17, where two main genetic substructures merged as components of the genetic pool of the Checkered Giant breed.

The three small/dwarf breeds (Ermine, Dwarf Lop and Coloured Dwarf) shared some common genetic ancestry. Subsequent interbreed stratification was probably able to distinguish more clearly the Ermine (a small body sized breed) and the Dwarf Lop genetic pools (as shown at higher Ks). Some heterogeneity remained in the Coloured Dwarf. As this breed is not fixed for any colour, it might have experienced several introgressions from other morphs increasing its genetic heterogeneity. From the genetic structures of these three breeds, it could be possible to speculate that the general complex and quantitative genetic features that determine the small body size might have been commonly shared by all these three breeds that then were distinguished by some monogenic or oligogenic features determining dwarfism in two of them (Dwarf Lop and Coloured Dwarf), which might further reduce their body size. The ear lop characteristic should

have also contributed to genetically separating the Dwarf Lop from the Coloured Dwarf. This is also evident from the analyses of the signature of selection obtained in this study and our previous study carried out in these two breeds (Ballan et al., 2022a). Different genomic regions containing candidate genes might be involved in determining different dwarfisms and the adapted skull structure requested by the ear lop phenotype (Carneiro et al., 2017; Ballan et al., 2022a; Fontanesi, 2021c).

The Rex breed, defined by its peculiar coat structure, showed some common ancestry with the two dwarf breeds at K = 11, suggesting a partially shared genetic background, which then diverged at higher Ks, according to a breed stratification process mainly led by the *Rex*^{1} locus genotype, that, according to the results of signature of selection analyses, we confirmed to be caused by variability at the *LIPH* gene, as previously reported (Diribarne et al., 2011). The Rex breed population analysed with a few microsatellites by Alves et al. (2015) showed some undefined substructures, probably derived by admixture with the New Zealand White breed, suggesting that the genetic material coming from different countries might have experienced some different introgression. As we did not include the New Zealand White breed, we could not confirm these potential genetic relationships.

As expected, the two silver breeds (Champagne d'Argent and Italian Silver) shared large common ancestral genetic background with some substructures in the Italian Silver breed probably derived by the breeding program aimed to improve performance traits in this male meat line. Several substructures were also evident in the other two meat breeds at all Ks, which can be attributed, again, to their breeding programs, which introduced the blood of different genetic origin to enlarge their genetic background and maximize the genetic progress.

Rhinelander and Belgian Hare breeds were very well differentiated from all other breeds. Belgian Hare resulted well separated from all other breeds also in the study of Alves et al. (2015), further suggesting a quite early demographic separation of this breed, as could be expected from the lean, long and thin body structure (very peculiar features compared to all other breeds) that characterizes Belgian Hare rabbits.

The complete picture of signatures of selection in the rabbit genome that we obtained in this study (by adding the results of the two haplotype-based approaches and combining this information with previous results obtained in the same rabbit breeds; Ballan et al., 2022a, 2022b), further confirm and extend the relevance of many genetic features for the genetic differentiation of groups of breeds and single breeds. Considering all breeds and all applied methods, signatures of selection were identified in a total of about 1.8 Gb of the rabbit genome. From this picture, it emerged that a large fraction of the genome experienced some relevant modifications in the process of breed constitution (considering that we analysed only a few breeds over the several tens or hundreds that have been constituted). On average, independent signals per breed covered about 113 Mb and if we considered only the most consistent results, again, on average, at least three methods identified 25 Mb, 8 Mb were identified by at least four methods and 1.2 Mb were identified by at least five methods. These highly consistent regions might contain very important genetic features that potentially should largely contribute to the specific genetic footprints of these breeds.

The role of the genes annotated in signatures of selection regions, known from other studies and other species, could be used to derive functional effects linked to some characteristics of the considered breeds or groups of breeds. However, more than one gene is usually annotated in this regions we identified. In some cases, it was possible to highlight the most relevant candidates according to their already well established roles. For example, all genes involved in pigmentation

could be easily highlighted, as most breeds and groups of breeds can be clearly distinguished through their coat colour and colour patterns. The list of these genes identified with the |iHS| and XP-EHH approaches included *ASIP*, *EDNRA*, *EDNRB*, *KIT*, *KITLG*, *MITF*, *OCA2*, *TYR* and *TYRP1* (Figures 3 and S1), some of which have also been reported using previously applied methods (Ballan et al., 2022a, 2022b), others were newly identified in this study. Therefore, the list of pigmentation relevant genes reported in this study and in the other studies that we carried out (Ballan et al., 2022a, 2022b) is quite large further indicating that coat colours and colour patterns are very important in establishing and then recognizing a rabbit breed (Fontanesi, 2021a, 2021b). Here is also interesting to mention the signature of selection identified in Champagne d'Argent on OCU9 in the correspondence of the *MITF* gene, which in humans is involved in the hair graying process (Harris et al., 2018) and melanocyte survival, migration, proliferation, and differentiation (Saleem et al., 2019). This region did not emerge in our previous study based on F_{ST} that however highlighted other genomic regions potentially involved in the greying phenotype (Ballan et al., 2022a). The *Silver* locus, inferred by classical genetic studies based on colour segregation, has been suggested to be the genetic factor underlying this peculiar progressive greying phenotype that characterizes this breed (Fontanesi, 2021b). However, the inconsistent results obtained with different signatures of selection approaches, that would indirectly aim to identify the genetic factors affecting the hair greying phenotype in this breed, might indicate that the genetic mechanisms determining this coat colour could be more complex than previously expected.

How we phenotypically grouped the breeds based on body size, which also reflects some common genetic origin (as also evidenced by the results of the ADMIXTURE analysis), indicated interesting patterns of signatures of selection that included some genes already known to affect

similar traits in other species. For example, *LCORL* has already been reported to affect body size and stature in humans and several other mammals (Price et al., 2011; Rubin et al., 2012; Signer et al., 2012; Bouwman et al., 2018; Plassais et al., 2019). Our previous study based on F_{ST} already showed a signature of selection including this gene but in small/dwarf breeds (Ballan et al., 2022a). A similar signature of selection in this gene region was also observed in dwarf rabbits by Carneiro et al. (2017). Here, we report signatures of selection including *LCORL* in Italian White rabbits and, in turn in all meat breeds. This result highlighted some putative alternative alleles at this gene region, probably affecting body structure in the opposite direction than that we can infer in the small/dwarf breeds, in line with what has been reported in humans and other mammals (Price et al., 2011; Rubin et al., 2012; Signer et al., 2012; Bouwman et al., 2018; Plassais et al., 2019). Similar bidirectional signatures of selection were identified in the *HMGA2* gene region that was highlighted in small/dwarf breeds and, with a putative opposite functional effect, also in giant breeds and in spotted breeds (that have large or medium body size). Variability in the *HMGA2* has been consistently associated with body size and stature in humans and several mammals (Weedon et al., 2007; Boyko et al., 2010; Makvandi-Nejad et al., 2012; Bouwman et al., 2018). A large deletion in this gene is the causative mutation of a form of dwarfism in rabbit and where only heterozygous animals at this mutated region are viable (Carneiro et al., 2017). In our previous study based on F_{ST} (Ballan et al., 2022a), we could only capture the signature of selection in the giant breeds but not the selection sweep that we identified here with the |iHS| values in the small breed Ermine and the Dwarf Lop breed (and in turn, in the group of small/dwarf breeds). Again, this result confirms the complementarity of the different methods we applied to identify signatures of selection in the rabbit genome. Many other regions, that emerged in the analyses carried out for

the small/dwarf and giant breeds may contain relevant genes involved in defining body size in rabbits, a complex and polygenic trait, as already demonstrated in all many other mammals.

Another signature of selection was identified with both |iHS| and XP-EHH methods in Checkered Giant on OCU9 in a region that includes the *PTPN2* gene. This gene region also strongly emerged in our previous study but as a main outlier in the meat breeds (Ballan et al., 2022a). This gene has been associated with Crohn's disease, a chronic inflammation involving the whole digestive tract (Festen et al., 2011). Therefore, *PTPN2* might have an important role in the functionality of the digestive structures. It is worth mentioning here that Checkered Giant rabbits are usually carriers of a genetic defect associated with the *English spotting* locus (probably derived by a mutation in the *KIT* gene), which when in homozygous condition it causes deleterious megacolon in the rabbits but that is usually silent in the heterozygous state (Fontanesi et al., 2014b). Therefore, it could be speculatively interesting to pair these two genetic elements in Checkered Giant: one at the *PTPN2* gene region, which could play a corrective function of a defective or semi-defective genotype at the second gene region, the *KIT*, to reduce the negative impact of the megacolon in the breed population. Further studies are needed to evaluate this potential epistatic role of *PTPN2* in this context.

The results of the functional enrichment analysis based on the genes included in the combined list of signatures of selection that we were able to obtain for the rabbit genome, even if derived from only 13 rabbit breeds, may indicate that the processes of breed constitution involved many genomic regions, which resulted in the modification of many different functions, with just very few major functional patterns for body size/shape (as derived by the GO term: axial length) and pigmentation processes (as derived by the GO term: eye color). Therefore, it seems that the second step of the domestication process of the rabbit, that can be considered when breeds started

to be constituted/defined, included a shift from the genes/functions of the first stage (genes mainly involved in behavioral and reproduction adaptation; Carneiro et al., 2014), to many other genes and functions with some major modifications that occurred for genes involved in the characterization of exterior traits (morphology and colour), that are mainly targeted by fancy breeders.

5. CONCLUSIONS

This study further demonstrated that high throughput genotyping data could be very useful to explore the population structure of rabbit breeds and to obtain information useful to reconstruct their genetic history and sometimes the complex series of admixture and introgression events that have shaped their genetic structures. It is also interesting to note that the use of different methods to detect signatures of selection highlighted important regions that could not emerge using just one or another approach. The signatures of selection left in the genome of rabbit genetic resources emerged to be at least partly consistent with their phenotypic similarities. Moreover, these selection sweeps derived by the genetic events and divergent artificial selection trajectories that have differentiated many genetic resources in the domestic rabbit, provide useful hints to go into more detail and better analyse some genes that could contribute to the genetic variability of external traits in this species and, as a model, also in other mammals. Other approaches can be added to what we reported in our studies to increase further and improve the obtained signature of the selection picture over the rabbit genome. Many other rabbit breeds remain to be investigated at the genome level. Their full characterization may identify additional hints able to explain the large phenotypic diversity in this domesticated species.

ACKNOWLEDGEMENTS

This study was funded by the PSRN (Programma di Sviluppo Rurale Nazionale) Cun-Fu and Cun-Fu 2 projects (co-funded by the European Agricultural Fund for Rural Development of the European Union and by the Italian Ministry of Agriculture, Food and Forestry - MASAF) and by the University of Bologna RFO 2019-2022 programs.

CONFLICT OF INTEREST

The authors declare they do not have any competing interests.

AUTHOR CONTRIBUTION

L.F. designed the study, interpreted the results, and obtained funding. L.F., F.B. and M.B. wrote the paper. M.B, S.B, F.B. and G.S. conducted data and bioinformatic analyses. M.S. and R.N. provided samples and data. F.B., S.B. and G.S. contributed to data interpretation. All authors read and approved the submitted version.

DATA AVAILABILITY

Genotyping data can be shared after the signature of an agreement on their use with the University of Bologna and ANCI. All requests should be addressed to luca.fontanesi@unibo.it.

REFERENCES

Aigner, B., Besenfelder, U., Müller, M., & Brem, G. (2000). Tyrosinase gene variants in different rabbit strains. *Mammalia Genome, 11*, 700–702. https://doi.org/10.1007/s003350010120

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Alves, J. M., Carneiro, M., Afonso, S., Lopes, S., Garreau, H., Boucher, S., & Ferrand, N. (2015). Levels and patterns of genetic diversity and population structure in domestic rabbits. *PLoS One*, *10*, e0144687. https://doi.org/10.1371/journal.pone.0144687

Ballan, M., Bovo, S., Schiavo, G., Schiavitto, M., Negrini, R., & Fontanesi, L. (2022a). Genomic diversity and signatures of selection in meat and fancy rabbit breeds assessed using high density SNP information. *Genetics Selection Evolution*, 54, 3. https://doi.org/10.1186/s12711-022-00696-9

Ballan, M., Schiavo, G., Bovo, S., Schiavitto, M., Negrini, R., Frabetti, A., Fornasini, D., & Fontanesi, L. (2022b). Comparative analysis of genomic inbreeding parameters and runs of homozygosity islands in several fancy and meat rabbit breeds. *Animal Genetics*, *53*, 849–862. https://doi.org/10.1111/age.13264

Bertolini, F., Moscatelli, G., Schiavo, G., Bovo, S., Ribani, A., Ballan, M., & Fontanesi, L. (2022). Signatures of selection are present in the genome of two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colours. *Journal of Animal Breeding and Genetics*, *139*, 307–319. https://doi.org/10.1111/jbg.12659

Bertolini, F., Schiavo, G., Bovo, S., Sardina, M. T., Mastrangelo, S., Dall'Olio, S., & Fontanesi, L. (2020). Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano-Reggiano cheese production system. A*nimal*, *14*, 921–932. https://doi.org/10.1017/S1751731119003318

Boucher, S., Garreau, H., Boettcher, P., & Bolet, G. (2021). Rabbit breeds and lines and genetic resources. In: Fontanesi L, editor. *The Genetics and Genomics of the Rabbit*, pp. 23-37. CAB International. Wallingford, UK. https://doi.org/10.1079/9781780643342.0002

Bovo, S., Schiavo, G., Utzeri, V. J., Ribani, A., Schiavitto, M., Buttazzoni, L., & Fontanesi, L. (2021). A genome-wide association study for the number of teats in European rabbits (*Oryctolagus cuniculus*) identifies several candidate genes affecting this trait. *Animal Genetics*, *52*, 237–243. https://doi.org/10.1111/age.13036

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., … Hayes, B. J. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics*, *50*, 362–367. https://doi.org/10.1038/s41588-018-0056-5

Boyko, A. R., Quignon, P., Li, L., Schoenebeck, J. J., Degenhardt, J. D., Lohmueller, K. E., … Ostrander, E. A. (2010). A simple genetic architecture underlies morphological variation in dogs. *PLoS Biology*, *8*, e1000451. https://doi.org/10.1371/journal.pbio.1000451

Callou, C. (2003). *De la garenne au clapier: Étude archéozoologique du lapin en Europe occidentale* (Vol. 189, pp. 356). Mémoires du Muséum national d'Histoire naturelle, 2-85653-545-3.

Carneiro, M., Rubin, C. J., Di Palma, F., Albert, F. W., Alföldi, J., Martinez Barrio, A., … Andersson, L. (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*, *345*, 1074–1079. https://doi.org/10.1126/science.1253714

Carneiro, M., Hu, D., Archer, J., Feng, C., Afonso, S., Chen, C., … Andersson, L. (2017). Dwarfism and altered craniofacial development in rabbits is caused by a 12.1 kb deletion at the *HMGA2* locus. *Genetics*, *205*, 955–965. https://doi.org/10.1534/genetics.116.196667

Carneiro, M., Vieillard, J., Andrade, P., Boucher, S., Afonso, S., Blanco-Aguiar, J. A., … Andersson, L. (2021). A loss-of-function mutation in RORB disrupts saltatorial locomotion in rabbits. *PLoS Genetics*, *17*, e1009429. https://doi.org/10.1371/journal.pgen.1009429

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, s13742–015–0047–8. https://doi.org/10.1186/s13742-015-0047-8

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 1-14. https://doi.org/10.1186/1471-2105-14-128

Delaneau, O., Marchini, J., & Zagury, J. F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, *9*, 179–181. https://doi.org/10.1038/nmeth.1785

Demars, J., Iannuccelli, N., Utzeri, V. J., Auvinet, G., Riquet, J., Fontanesi, L., & Allain, D. (2018). New insights into the melanophilin (*MLPH*) gene affecting coat color dilution in rabbits. *Genes*, *9*, 430. https://doi.org/10.3390/genes9090430

Diribarne, M., Mata, X., Chantry-Darmon, C., Vaiman, A., Auvinet, G., Bouet, S., … Guérin, G. (2011). A deletion in exon 9 of the *LIPH* gene is responsible for the rex hair coat phenotype in rabbits (*Oryctolagus cuniculus*). *PLoS One*, *6*, e19281. https://doi.org/10.1371/journal.pone.0019281

Festen, E. A., Goyette, P., Green, T., Boucher, G., Beauchamp, C., Trynka, G., … Rioux, J. D. (2011). A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2,

TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. *PLoS Genetics, 7,* e1001283. https://doi.org/10.1371/journal.pgen.1001283

Fontanesi, L. (2021a). Rabbit genetic resources can provide several animal models to explain at the genetic level the diversity of morphological and physiological relevant traits. *Applied Sciences, 11,* 373. https://doi.org/10.3390/app11010373

Fontanesi, L. (2021b) Genetics and molecular genetics of coat colour in the European rabbit. In: Fontanesi L, editor. *The Genetics and Genomics of the Rabbit,* pp. 38–42. CAB International. Wallingford, UK. https://doi.org/10.1079/9781780643342.0006

Fontanesi, L. (2021c) Genetics and molecular genetics of morphological and physiological traits and inherited disorders in the European rabbit. In: Fontanesi L, editor. *The Genetics and Genomics of the Rabbit,* pp. 120–162. CAB International. Wallingford, UK. https://doi.org/10.1079/9781780643342.000

Fontanesi, L., Forestier, L., Allain, D., Scotti, E., Beretti, F., Deretz-Picoulet, S., … Oulmouden, A. (2010a). Characterization of the rabbit agouti signaling protein (*ASIP*) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. *Genomics, 95,* 166–175. https://doi.org/10.1016/j.ygeno.2009.11.003

Fontanesi, L., Scotti, E., Allain, D., & Dall'Olio, S. (2014a). A frameshift mutation in the melanophilin gene causes the dilute coat colour in rabbit (*Oryctolagus cuniculus*) breeds. *Animal Genetics, 45,* 248–255. https://doi.org/10.1111/age.12104

Fontanesi, L., Scotti, E., Colombo, M., Beretti, F., Forestier, L., Dall'Olio, S., … Oulmouden, A. (2010b). A composite six bp in-frame deletion in the melanocortin 1 receptor (*MC1R*) gene is associated with the Japanese brindling coat colour in rabbits (*Oryctolagus cuniculus*). *BMC Genetics, 11,* 59. https://doi.org/10.1186/1471-2156-11-59

Fontanesi, L., Tazzoli, M., Beretti, F., & Russo, V. (2006). Mutations in the melanocortin 1 receptor (*MC1R*) gene are associated with coat colours in the domestic rabbit (*Oryctolagus cuniculus*). *Animal Genetics*, *37*, 489–493. https://doi.org/10.1111/j.1365-2052.2006.01494.x

Fontanesi, L., Utzeri, V. J., & Ribani, A. (2021). The evolution, domestication and world distribution of the European rabbit (*Oryctolagus cuniculus*). In: Fontanesi L, editor. *The Genetics and Genomics of the Rabbit*, pp. 1–22. CAB International. Wallingford, UK. https://doi.org/10.1079/9781780643342.0001

Fontanesi, L., Vargiolu, M., Scotti, E., Latorre, R., Faussone Pellegrini, M. S., Mazzoni, M., … De Giorgio, R. (2014b). The *KIT* gene is associated with the *English spotting* coat color locus and congenital megacolon in Checkered Giant rabbits (*Oryctolagus cuniculus*). *PLoS One*, *9*, e93750. https://doi.org/10.1371/journal.pone.0093750

Gautier, M., & Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, *20*, 3128–3143. https://doi.org/10.1111/j.1365-294X.2011.05163.x

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, *17*, 78–90. https://doi.org/10.1111/1755-0998.12634

González-Rodríguez, A., Munilla, S., Mouresan, E. F., Cañas-Álvarez, J. J., Díaz, C., Piedrafita, J., … Varona, L. (2016). On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. *Genetics Selection Evolution*, *48*, 81. https://doi.org/10.1186/s12711-016-0258-1

Harris, M. L., Fufa, T. D., Palmer, J. W., Joshi, S. S., Larson, D. M., Incao, A., … Pavan, W. J. (2018). A direct link between MITF, innate immunity, and hair graying. *PLoS Biology*, *16*, e2003648. https://doi.org/10.1371/journal.pbio.2003648

Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., & Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, *15*, 246. https://doi.org/10.1186/1471-2164-15-246

Lehner, S., Gähle, M., Dierks, C., Stelter, R., Gerber, J., Brehm, R., & Distl, O. (2013). Two-exon skipping within *MLPH* is associated with coat color dilution in rabbits. *PLoS One*, *8*, e84525. https://doi.org/10.1371/journal.pone.0084525

Letko, A., Ammann, B., Jagannathan, V., Henkel, J., Leuthard, F., Schelling, C., Carneiro, M., Drögemüller, C., & Leeb, T. (2020). A deletion spanning the promoter and first exon of the hair cycle-specific *ASIP* transcript isoform in black and tan rabbits. *Animal Genetics*, *51*, 137–140. https://doi.org/10.1111/age.12881

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*, 67–77. https://doi.org/10.1111/1755-0998.12592

Maiorano, A. M., Lourenco, D. L., Tsuruta, S., Ospina, A. M. T., Stafuzza, N. B., Masuda, Y., … Silva, J. A. I. D. V. (2018). Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. *PLoS One*, *13*, e0200694. https://doi.org/10.1371/journal.pone.0200694

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., … Sutter, N. B. (2012). Four loci explain 83% of size variation in the horse. *PLoS One*, *7*, e39929. https://doi.org/10.1371/journal.pone.0039929

Plassais, J., Kim, J., Davis, B. W., Karyadi, D. M., Hogan, A. N., Harris, A. C., … Ostrander, E. A. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature Communications, 10,* 1489. https://doi.org/10.1038/s41467-019-09373-w

Pryce, J. E., Hayes, B. J., Bolormaa, S., & Goddard, M. E. (2011). Polymorphic regions affecting human height also control stature in cattle. *Genetics, 187,* 981–984. https://doi.org/10.1534/genetics.110.12394

Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science, 166,* 133–143. https://doi.org/10.1016/j.livsci.2014.05.003

Rubin, C. J., Megens, H. J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, … Andersson, L. (2012). Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the USA, 109,* 19529–19536. https://doi.org/10.1073/pnas.1217149109

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., … Lander, E. S. (2006). Positive natural selection in the human lineage. *Science, 312,* 1614–1620. https://doi.org/10.1126/science.1124309

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature, 449,* 913–918. https://doi.org/10.1038/nature06250

Saleem, M. D. (2019). Biology of human melanocyte development, Piebaldism, and Waardenburg syndrome. *Pediatric dermatology, 36*(1), 72-84. https://doi.org/10.1111/pde.13713

Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., & Mishra, B. P. (2020). Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livestock Science, 241*, 104257. https://doi.org/10.1016/j.livsci.2020.104257

Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T., & Rieder, S. (2012). A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One*, *7*, e37282. https://doi.org/10.1371/journal.pone.0037282

Ueno, H., Huang, X., Tanaka, Y., & Hirokawa, N. (2011). KIF16B/Rab14 molecular motor complex is critical for early embryonic development by transporting FGF receptor. *Developmental Cell*, *20*, 60–71. https://doi.org/10.1016/j.devcel.2010.11.008

Utsunomiya, Y. T., Pérez O'Brien, A. M., Sonstegard, T. S., Van Tassell, C. P., do Carmo, A. S., Meszaros, G., Sölkner, J., & Garcia, J. F. (2013). Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS One*, *8*, e64280. https://doi.org/10.1371/journal.pone.0064280

Utzeri, V. J., Ribani, A., & Fontanesi, L. (2014). A premature stop codon in the TYRP 1 gene is associated with brown coat colour in the European rabbit (Oryctolagus cuniculus). *Animal Genetics*, *45*, 600–603. https://doi.org/10.1111/age.12171

Utzeri, V. J., Ribani, A., Schiavo, G., & Fontanesi, L. (2021). Describing variability in the tyrosinase (*TYR*) gene, the *albino* coat colour locus, in domestic and wild European rabbits. *Italian Journal of Animal Science*, *20*, 181–187. https://doi.org/10.1080/1828051X.2021.1877574

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*, e72. https://doi.org/10.1371/journal.pbio.0040072

Weedon, M. N., Lettre, G., Freathy, R. M., Lindgren, C. M., Voight, B. F., Perry, J. R., … Frayling, T. M. (2007). A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genetics, 39,* 1245–1250. https://doi.org/10.1038/ng2121

Whitman B.D. (2004) *Domestic Rabbits & Their Histories: Breeds of the World*. Leathers Publishing. Overland Parks, KS, USA.

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., … Lee, J. Y. (2022). A saturated map of common genetic variants associated with human height. *Nature, 610,* 704–712. https://doi.org/10.1038/s41586-022-05275-y

Zeder, M. A. (2012). The domestication of animals. *Journal of Anthropological Research, 68,* 161–190.

Zeuner F.E. (1963) A History of Domesticated Animals. Harper & Row Publishers, Evanston, UK

Figures

Figure 1. Multidimensional scaling plots. The first three components are provided.

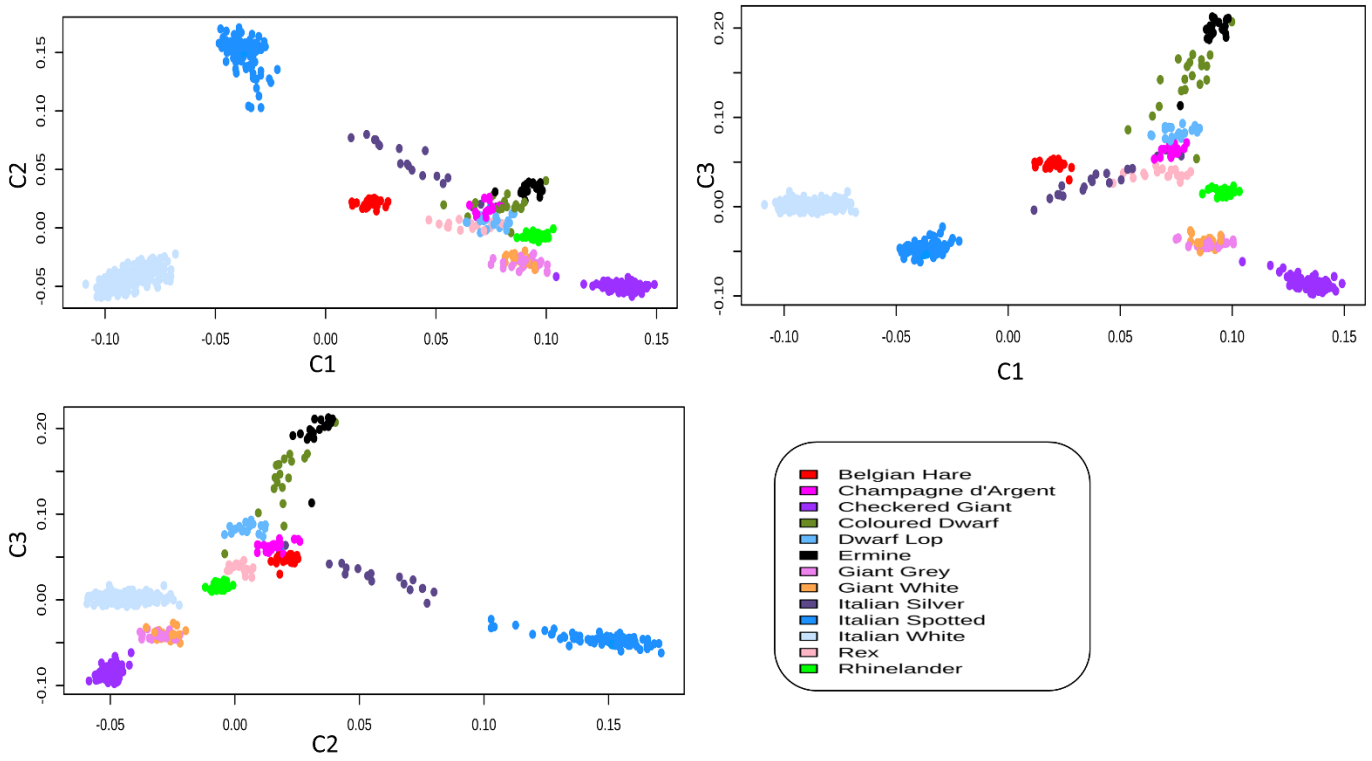


Figure 2. Results of the ADMIXTURE analysis. a) Cross validation (CV) errors with K from 1 to 17. b) Plot distribution with K = 11, 13 and 17 of the averaged admixture ancestries per breed. Putative subpopulations are labelled with a different colour.

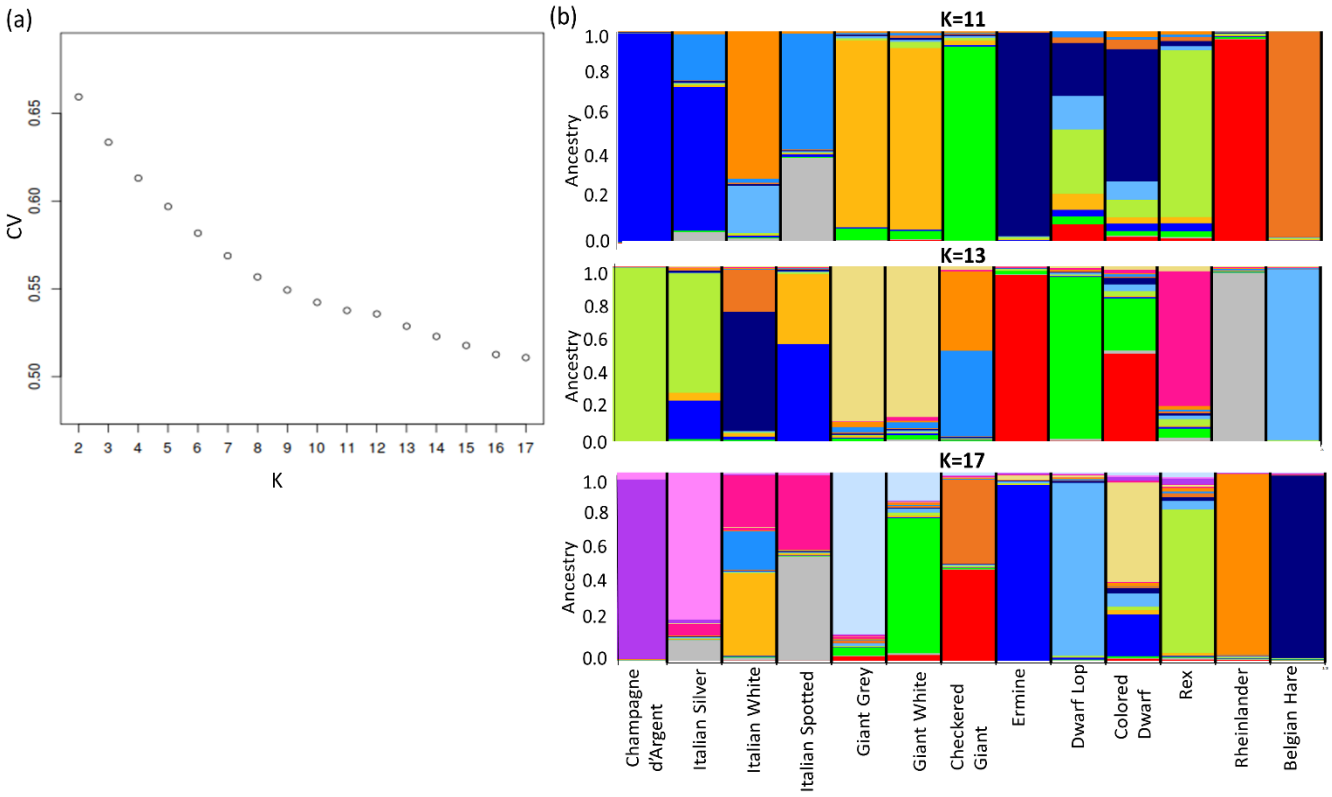


Figure 3. The Manhattan plots obtained for the breeds grouped according to their coat colour, body size and use (i.e. meat breeds). (a) Manhattan plots obtained with |iHS| values. (b) Manhattan plots obtained with XP-EHH values.

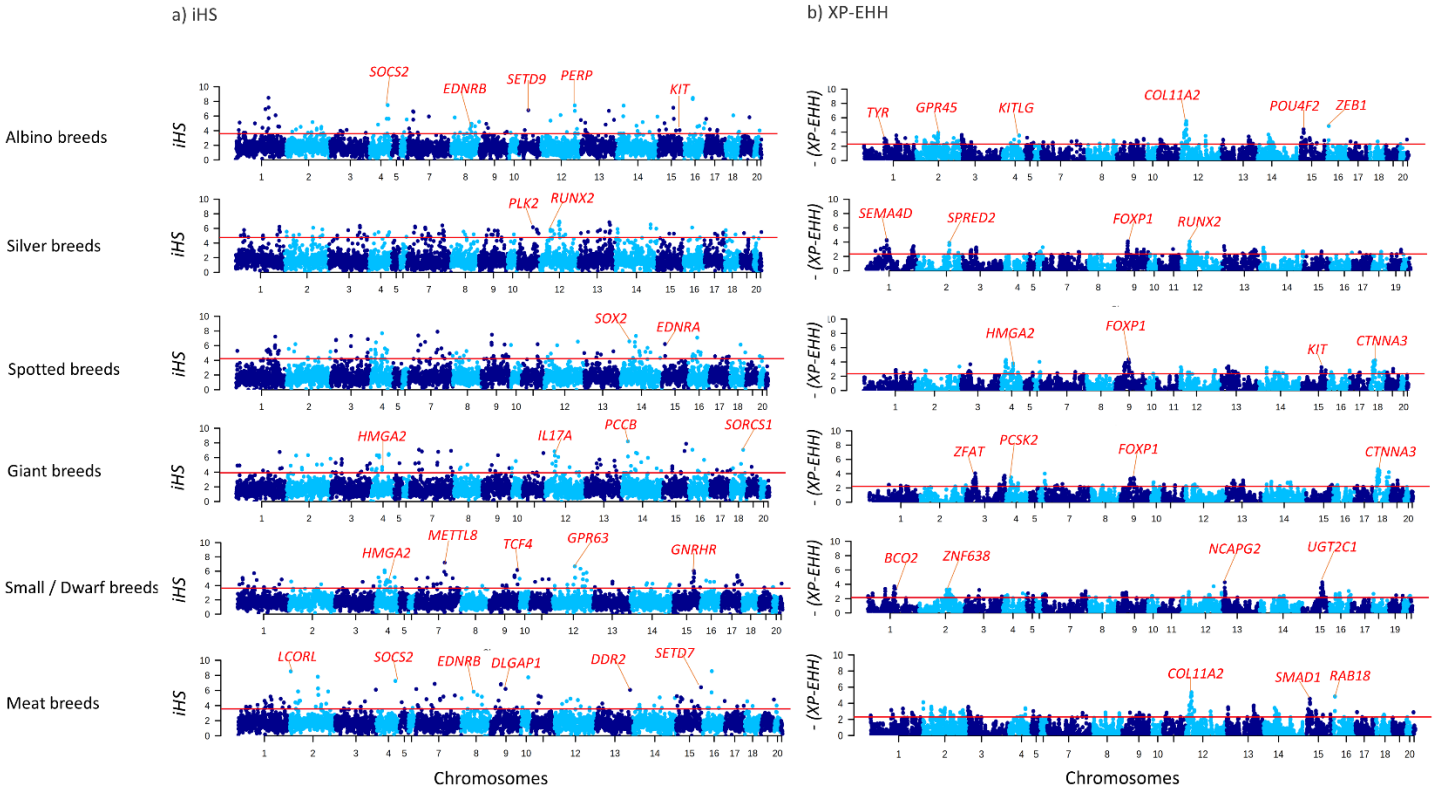


Figure 4. An overview of the genomic regions covered by signatures of selection identified by three, four or five different methods. Detailed information is reported in Supplementary material.

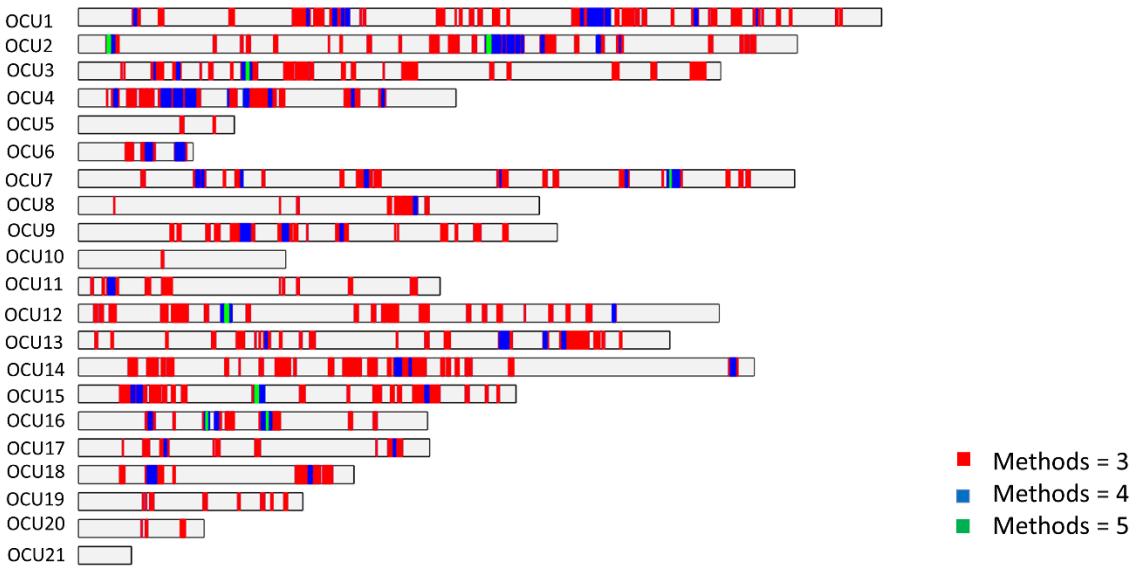


Table 1. List of the top 20 integrated haplotype score measures |iHS| for the 350 kb genomic windows and their closest genes within flanked genomic region of 500 kb, with information on the *Oryctolagus cuniculus* chromosome (OCU) position, identified with the single breed analysis or the group of breeds analyses. The full information of |iHS| for single breed approach and groups of breeds approach are presented in Table S2 and Table S3 respectively.

|iHS|	**OCU:position**	**Breeds/groups of breeds[1]**	**Closest genes**	**Distance (kb)[2]**
7.408	1:122550000-123500000	Italian White	*UBE2I;TAF1D;PANX1;VSTM5;MED17;C11orf54;SMCO4;HEPHL1;CEP295;CCDC67*	100
7.235	1:149500000-150450000	Spotted breeds	*NLRP10;EIF3F;LMO1;TRIM66;TUB;STK33;RIC3*	269
7.821	2:107150000-108100000	Meat production	*REG3G*	142
8.684	2:8450000-9400000	Italian White, Meat production	*LCORL*	45
7.344	4:43800000-44750000	Checkered Giant	*WIF1;MSRB3;HMGA2;LEMD3*	100
7.688	4:46950000-47900000	Spotted breeds	*IFNG;IL26;IL22;RAP1B;SLC35E3;MDM2;MDM1;NUP107;CPM*	162
7.58	4:56400000-57350000	Italian White	*NAV3*	200
7.495	4:72850000-73800000	Albino breeds	*LOC108176670;SOCS2;CRADD;PLXNC1;CEP83*	500
7.266	4:72850000-73800000	Meat production	*SOCS2;CRADD;PLXNC1;CEP83*	400
7.386	7:38550000-39500000	Spotted breeds	*PTPN12;LOC108176821;FGL2;LRRC17;ARMC10;GSAP;CCDC146;FAM185A;FBXL13*	110
7.219	7:47650000-48600000	Giant Grey	*LOC100355039;LRRN3;IMMP2L*	300

7.493	9:42400000-43350000	Spotted breeds	*ZNF717*	128
7.407	9:49750000-50700000	Checkered Giant	*CHMP1B;AFG3L2;CIDEA;MPPE1;LOC108177123;TUBB6;IMPA2;GNAL;LOC108177122*	73
8.472	12:128500000-129450000	Italian White	*HEBP2;PERP;CCDC28A;ARFGEF3;ECT2L;REPS1;NHSL1*	16
7.448	12:128500000-129450000	Albino breeds	*HEBP2;LOC108177357;PERP;CCDC28A;ARFGEF3;ECT2L;REPS1;NHSL1*	500
8.209	14:29100000-30050000	Giant breeds	*MSL2;PCCB;EPHB1;PPP2R3A*	98
7.426	14:30500000-31450000	Albino breeds	*SLC35G2;IL20RB;NCK1*	107
7.948	15:64800000-65750000	Italian White	*ARHGAP24*	500
8.558	16:34000000-34950000	Meat production	*LOC100345607;KCNK1;PCNX2*	495
8.323	16:34700000-35650000	Albino breeds	*MAP10;LOC100345607;NTPCR;PCNX2;SIPA1L2*	0

[1] The breed or group of breeds listed first are reported according to the top |iHS|. Other breeds or groups of breeds are reported when the same region was evidenced over the thresholds in other populations.

[2] The distance between the genomic window and the most relevant or closest gene in kb.

Table 2. List of all regions identified with at least four methods in the single breed analysis [|iHS| and XP-EHH approaches were reported in this study; the results obtained using F_{ST}, *PCAdap*t and ROH island (ROH) approaches have been obtained from Ballan et al. (2022a, 2022b)].

Genomic region[1]	Methods	Breeds	Annotated genes[2]	Trait[3]	Reference[4]
1:127150001-127200000	|iHS|, XP-EHH, *PCAdapt*, F_{ST}	Italian White	*CHORDC1, TRIM77, NOX4,* ***TYR***	Coat color	Aigner et al. (2000)
3:41000001-41050000	ROH, |iHS|, XP-EHH, PCAdapt, F_{ST}	Giant White	*ADRA1B, TTC1, UBLCP1, IL12B,* ***EBF1****, PWWP2A, RNF145*	Number of teats in rabbits	Bovo et al. (2021)
4:20050001-20100000	|iHS|, XP-EHH, *PCAdapt*, F_{ST}	Dwarf Lop	*KIF16B, MACROD2*	-	-
4:40300001-40350000	|iHS|, XP-EHH, *PCAdapt*, F_{ST}	Checkered Giant	*APOF, ANKRD52, NACA, FAM19A2, STAT2, IL23A, PAN2, PTGES3, RBMS2, TIMELESS, COQ10A, SPRYD4, CNPY2, ATP5B, BAZ2A;MIP, NABP2, SLC39A5, GLS2*	-	-
6:18000001-18050000	|iHS|, XP-EHH, *PCAdapt*, F_{ST}	Italian Silver	*ASPHD1, CORO1A, TBX6, CDIPT, YPEL3, XPO6, SLX1A, PPP4C, SGF29, MAPK3, BOLA2B, DOC2A, KCTD13, IL4R, KIAA0556, MVP, GSG1L, ALDOA, INO80E, SEZ6L2, TAOK2, FAM57B, IL21R, GTF3C1,*	-	-

7:102350001-102400000	ROH, |iHS|; XP-EHH, F_{ST}	Giant White	*PRRT2, GDPD3, TMEM219, PAGR1, HIRIP3 FIGN, TRNAF-GAA, KCNH7*	-	-
9:38800001-38850000	|iHS|, XP-EHH, PCAdapt, F_{ST}	Belgian Hare	*EIF4E3, FOXP1, PROK2*	-	-
9:49650001-49700000	|iHS|, XP-EHH, *PCAdapt*, F_{ST}	Checkered Giant	*SEH1L, TUBB6, SPIRE1, ANKRD30B, CEP76, AFG3L2, IMPA2, PSMG2,* ***PTPN2****, GNAL, CHMP1B, PRELID3A, MPPE1, CIDEA*	Crohn's disease	Festen et al. (2011)
9:63250001-63300000	|iHS|, XP-EHH, PCAdapt, F_{ST}	Checkered Giant	*SNRPD1, ESCO1, TRNAK-CUU, MIB1, GATA6, ABHD3, GREBs1L*	-	-
11:50150001-50200000	ROH, |iHS|, *PCAdapt*, F_{ST}	Belgian Hare	*CDH9*	-	-
15:80600001-80650000	ROH, |iHS|, XP-EHH, F_{ST}	Coloured Dwarf	*UGT2C1, UGT2B14, UGT2B16, UGT2B13, UGT2A1*	-	-
17:78450001-78500000	ROH, |iHS|, XP-EHH, F_{ST}	Checkered Giant	*GABRG3,* ***OCA2***	Coat color	Ballan et al. (2022a, 2022b)

[1] *Oryctolagus cuniculus* chromosome (OCU) and chromosome positions in OryCun2.0 genome version.

[2] Gene included in the genome regions. In bold, candidate genes with described function related to some external and functional traits.

[3] Traits that might be affected by variability in the genes indicated in bold in the previous column.

[4] References related to the involvement of the genes indicated in bold in the mentioned traits.

General conclusions

Genomic applications serve a pivotal role in the identification of deleterious alleles in livestock populations, aiming at the development of more robust and healthier breeding strategies. In parallel, these genomic tools could be utilized in the detection of signatures of selection, facilitating the improvement of breeding programs to incorporate candidates' genes for coat color, body size and meet production to enhance desired traits.

In this work, we first focused on the identification of deleterious DNA markers (SNPs and Haplotypes) from three Italian heavy pig breeds (Italian Large White, Italian Landrace, and Italian Duroc). The dataset includes around 10,000 samples genotyped using a 60K SNP bead chip. Deleterious alleles are usually identified by screening the analysed populations for DNA markers with homozygosity deficiency and deviation from Hardy-Weinberg equilibrium. The availability of the genotyped dataset for the DNA markers across three pig breeds allowed us to develop a genotype breed comparison approach. This approach discarded deleterious alleles presented in more than one pig breed assuming those deleterious alleles are breed-specific.

Haplotype construction was accomplished using two distinct tools, Beagle and SHAPIET2, haplotypes estimated by both tools were considered for the detection of deleterious haplotypes. This approach allowed us to discard the false positives associated with phased haplotypes, as each tool exhibited a certain degree of inaccuracy in estimating haplotypes from the genotyped dataset. The carrier frequencies of identified deleterious DNA markers are 7% to 17% for SNPs and 3 to 6% for haplotypes across the three pig populations. Interestingly, we confirmed that the haplotype-based approach is more effective in identifying deleterious alleles with lower frequency which might remain untagged by the genotyping tool.

A key limitation of this analysis was the lack of availability of a whole-genome sequencing dataset for individuals carrying the identified deleterious DNA markers. By calling the genomic variants from nine whole-genome sequencing samples carrying identified deleterious SNPs, we identified loss-of-function variants like frameshift, start lost, and splice donor mutations. These variants are in close to important candidate genes such as *IGF2BP1*, *ADGRL4*, *POMT2*, and *HGF*, suggesting a potential association between these deleterious alleles and critical genetic loci. According to the GWAS analysis of reproductive traits, the most significant deleterious DNA marker is HAP3, exhibiting a p-value of 0.001 for the total number of born alive traits. The beta value for this haplotype is -1.453, indicating an association with a reduction in the number of live-born piglets within the Italian Large White breed

In the future, there is a need to expand the sequencing efforts to encompass a larger cohort of carriers of the identified deleterious SNPs and haplotypes aiming to clarify the functional consequences and identify potential causal genetic variants. Furthermore, the integration of more extensive genotyped datasets, coupled with the collection of accurate phenotypic and pedigree information, will greatly contribute to the effective management of deleterious DNA markers within the analysed pig breeds.

The study detected 5,079 independent genomic regions under a selection of signatures across 13 rabbit breeds, spanning approximately 65% of the rabbit genome. These results revealed that a substantial portion of the rabbit genome experienced some relevant modifications in the processes of breed constitutions, considering that only a few breeds were included in the study over the several tens or hundreds that have been constituted. The identified genomic regions under selection of signatures included novel genes described for the first time in rabbits (e.g., *EDNRA*, *KITLG*,

OCA2 for pigmentation, *COL2A1*, *COL11A1*, *HOXD* for body size) alongside known genes described before in rabbits (e.g., *ASIP*, *MC1R*, *TYR* for pigmentation, *LIPH* for coat structure, *HMGA2* for body size). This represents the first comprehensive study in rabbits including diverse breeds and employing five different selection identification methods (Fst, ROH, *PCAdapt*, iHS, and the XP-EHH).

Despite the interesting results gained from our study, there are limitations to consider. Differentiating genomic regions between rabbit breeds may not always indicate selection due to complex breed history. Annotated gene functions can assist interpretation, but incomplete gene knowledge introduces uncertainties, especially for complex traits involving multiple genes. Enrichment analysis that have been performed did not provide clear insights, possibly due to the involvement of numerous genes in the genomic regions. Additionally, the study revealed selection signals in unassigned genomic scaffolds, emphasizing the need to enhance the quality of the rabbit reference genome assembly to more effectively uncover signals of selection at the chromosomal level, which may have been fragmented or overlooked in the existing investigation.

We still need to perform further analysis with additional breeds and populations and include more statistical approaches to expand this analysis of signatures of selection in the rabbit genome. Moreover, WGS data can be integrated with other 'omics' data, such as transcriptomics and proteomics, to gain a deeper understanding of the molecular mechanisms underlying selection signatures in the rabbit genome.